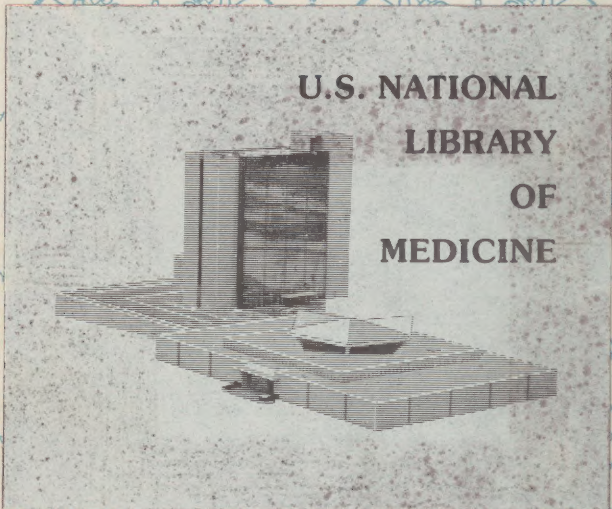

U.S. NATIONAL
LIBRARY
OF
MEDICINE

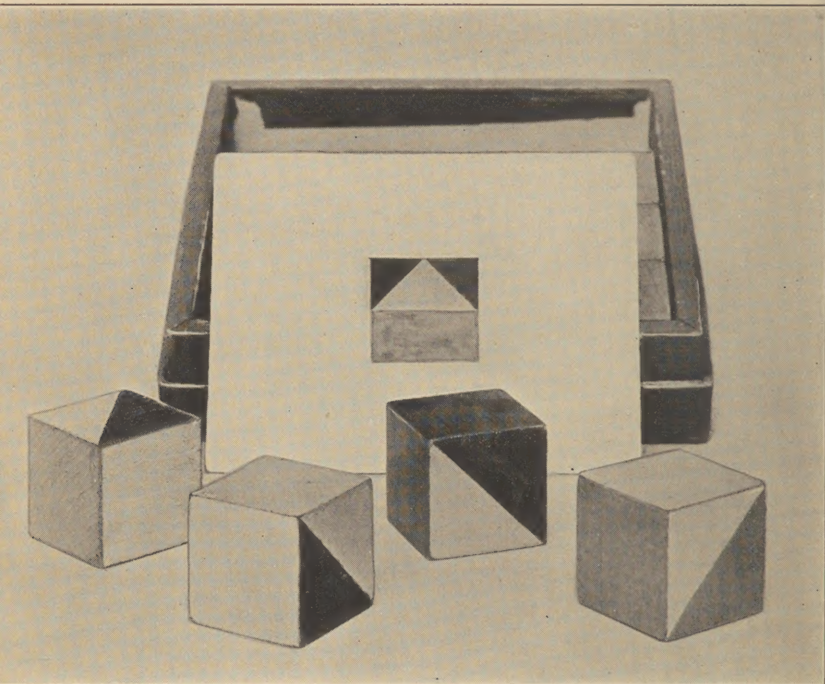
INTELLIGENCE MEASUREMENT

A Psychological and Statistical Study Based upon the Block-design Tests.

THE MACMILLAN COMPANY
NEW YORK · BOSTON · CHICAGO · DALLAS
ATLANTA · SAN FRANCISCO

MACMILLAN & CO., LIMITED
LONDON · BOMBAY · CALCUTTA
MELBOURNE

THE MACMILLAN CO. OF CANADA, LTD.
TORONTO



Courtesy Psychological Review Co.

THE BLOCK-DESIGN TEST

INTELLIGENCE MEASUREMENT

A Psychological and Statistical Study Based upon the Block-design Tests

BY

S. C. KOHS, Ph.D.

PSYCHOLOGIST, COURT OF DOMESTIC RELATIONS, PORTLAND, OREGON; ASSISTANT PROFESSOR OF PSYCHOLOGY, REED COLLEGE, AND PROFESSOR OF PSYCHOLOGY, PORTLAND CENTER, UNIVERSITY OF OREGON

New York
THE MACMILLAN COMPANY
1923

PRINTED IN THE UNITED STATES OF AMERICA

Set up and electrotyped. Published April, 1923.

Press of
J. J. Little & Ives Company
New York, U. S. A.

TO

HENRY HERBERT GODDARD

WHOSE PIONEER INTEREST IN INTELLIGENCE MEASUREMENT GREATLY ACCELERATED THE RAPID PROGRESS OF THIS PROMISING BRANCH OF SCIENCE, A MAN OF BROAD SYMPATHY, WITH KEEN UNDERSTANDING OF THE PROBLEMS CONFRONTING THOSE OF DEFICIENT MENTALITY, ONE TO WHOM THE WRITER OWES HIS INITIAL DIRECTION IN THIS BROADLY HUMANITARIAN FIELD,

THIS BOOK IS
GRATEFULLY
DEDICATED

PREFACE

New mental test methods of demonstrated validity are always welcome, and this is especially true just now of those which are relatively independent of the language factor. Other things being equal, any psychologist prefers to rate the intelligence of his subjects by what they can do rather than by what they say. Very often the "other things" are not equal, and the test which uses language possesses certain advantages that are not to be lightly set aside. The carefully worked out performance test, however, has a wide field of usefulness, as all who have engaged in clinical work will readily admit. The need for such tests has never been adequately met, largely because really valid ones are much more difficult to devise than is the case with tests which utilize language. In the upper ranges of intelligence, especially, most performance tests have but little differentiating value, simply because they do not draw heavily enough upon the higher mental processes. In the Block-design test, Dr. Kohs has largely overcome this difficulty, and has given us an intelligence scale which is certain to prove extremely useful. His success, we believe, is the direct result of his thoroughgoing appreciation of the real nature of the mental processes involved in what we call intelligence.

LEWIS M. TERMAN.

INTRODUCTION

"What is the nature of mind?" This question, propounded when man first became conscious of himself, still remains but inadequately answered. Nevertheless, some evident progress has been made. The light of scientific progress is gradually penetrating the various nooks and recesses of our mental life, and the machinery of thought, although dimly revealed, is becoming more apparent in its operation. This monograph is but a mere fragment, explaining little if anything regarding the dynamics of mental process, yet it hopes to place a variety of perplexing problems in a new perspective. We shall have occasion, for example, to examine some current definitions of intelligence, and we shall indicate wherein our research into the value and significance of completion and combination tests forces us to a reconsideration of the criteria of intelligence and to a restatement and a redefining of some fundamental principles. By mentioning in this connection the great importance of the "analytic-synthetic" tendency characteristic of all mental behavior, from the simplest sensory, ideational, and affective states to the most complex, we are really anticipating a later discussion of these matters. Suffice it to say that psychology has not yet interpreted our mental constructs in terms of a synthetic principle, although our philosophers have already blazed a somewhat ambiguous trail.

It has often been urged that the position of neutrality is one full of danger for its claimant. Although this may be adjudged a "behaviorist" study, nevertheless full weight has been given to those contributions of "structuralists" and "functionalists" which have thrown light upon our particular problem. The psychological family is at discord unless each member is willing, not alone to see the others' point of view with tolerance, but to accept at its real value any scientific contribution no matter what the standpoint of the experimenter. This is no apology for the present study — the remarks are intended merely to emphasize our community of problems, and the possibility that each school of method may throw some light on all of them. We are all confronted with the question, "What is the nature of mind?" Knowing the futility of any complete reply, we have accustomed ourselves either to ignore the question or to maneuver a Freudian "escape from reality" by losing ourselves in metaphysical speculation. If this piece of work yields but a little toward an additional understanding of this baffling problem our efforts will, indeed, be well rewarded.

S. C. Kohs.

Portland, Oregon,
August 30, 1921

CONTENTS

PART C: APPENDICES

PART A

THE PSYCHOLOGY OF ANALYSIS AND SYNTHESIS

THE PSYCHOLOGY OF ANALYSIS AND SYNTHESIS

The Place of Synthetic and Analytic Functioning in Psychic Activity

I. PHILOSOPHICAL ORIENTATION

The psychology of combinative or completive ability leads us back through the early history of psychology to the time when philosophers psychologized and psychologists philosophized. The limited confines of our study compel brevity, and we need go back no farther than the "associationist" school.

Hobbes, a staunch materialist, the contemporary of Bacon, Descartes, and Galileo, may rightly be called the father of "associational" psychology. Hobbes "materialized" mental activity. Psychic behavior was interpreted as mental physics or chemistry. All knowledge arises through sensations, which, combined in certain definite forms, yield the various complex mental processes. Conscious states are due to the refined movements of mental atoms controlled by the laws of association, the foundation-stones of mental functioning. Words are nothing but mere symbols, markers. It is their combination, their addition, their subtraction, which yields thought.

The legitimate successor of Bacon and Hobbes was John Locke. From a psychological point of view, it may be of interest to note that his "Essay on the Human Understanding" grew out of his attempt to answer the question: "What is the nature of sensations?" He began by asserting that there are no "innate ideas," in opposition to the position of Descartes and to the later views of Spinoza and Leibniz. At the beginning, mind is a "white piece of paper," upon which impressions are marked. Ultimately "nothing is in the intellect that has not been first in the sense." But the mind is not a passive receptacle. It can manipulate and combine ideas. These ideas are derivates of "simple ideas" out of which the more complex have been compounded. This compounding is explainable in terms of the laws of association.

Hume, an outstanding sensationalist, maintained that no ideas existed which were not derived from impressions, a position to which many of our present-day psychologists would claim adherence. Impressions are of two kinds: sensational (outer) and affective (inner). And ideas are copies or echoes of these simple elemental states. Much of his psychology is devoted to a discussion of the compounding or the combination of ideas, and this to such a degree, that his whole psychology becomes warped in its exaggeration of the actual part played by associational processes. The differences between conception, memory, imagination, judgment, and so on, are largely differences in the forms by means of which ideas combine. All mental behavior is just association between ideas. Ideas possess "forces" or "de-

terminations" which tend to bring and keep them together. The individual then is helpless, since these "forces" follow fixed, mechanical laws.

Leaving the "associationists," we find Kant, among philosophers, giving perhaps the clearest exposition of the hierarchies of mental functioning, progressing through hierarchies of mental "synthesis."

Kant, perhaps, has dealt more extensively with "mental synthesis" than any other single philosopher. His task as a philosopher was largely epistemological. His metaphysical structure, although promised, never appeared complete. It remained for others, Fichte, Hegel, Schelling, Herbart, and Schopenhauer, to take the step beyond. Kant was interested in the processes which give us knowledge. He postulated a threefold world: one, of subjective states (inner consciousness), another of phenomena (knowledge), and another of absolute reality (*Ding-an-Sich*). The first realm each individual can alone experience, the last no one can experience, the second, however, is common property and can be intelligently discussed by all of us. To Kant, subjective states are synthetic products, different in every one of us. And in each of his three books which compose his "Critique of Pure Reason" he discusses the different levels of synthetic activity: (*a*) in perception, (*b*) in understanding, (*c*) in reason. In his search for the processes which yield knowledge, he concluded that the important factor was the synthetic power of mind. In this sense mental functioning is dynamic. If one can imagine an individual who lacks this synthetic power, the ticking of a clock would be nothing more than tock, tock, tock.

To those, however, who do possess it, the clock's ticking is perceived as one, two, three, etc. The experience is cumulative, each item including its predecessors.

Kant very clearly differentiated between mental end-results and synthesis — one a product, the other a process. Synthesis is not subject to volitional vicissitudes, but is a heritage common to all. It is for that reason that we can agree upon such matters as physical laws, time-estimates, standards, and so on. Racial differences in thought and knowledge are differences largely of synthetic potentiality. Groups, therefore, may misunderstand each other because of this basic difference, since differences in synthetic power will lead to differences in thought and behavior. "The world is my representation," says Kant. Knowledge is a combination of sensations or ideas, "a unification of the manifold." And the expression of knowledge is through the judgment, which may be analytic or synthetic. Kant discusses quite extensively the distinction between the two types, but there is no need of our following his argument farther.

Few, if any, of the terms we use to express our thoughts come to us as an immutable, unchanged, and unchangeable heritage. Although we continue using the same terms, these take on changing meanings with the changes in our notions regarding the character of the individual and of the cosmos. Therefore, although we shall frequently utilize the terms "analysis" and "synthesis" throughout this monograph, they are not to be construed other than as here defined.

1. By "synthesis" we mean, on the one hand, the intensity of that fundamental force or condition innate in nervous protoplasm which binds neurons into complex systems, and, on the other hand, the capacity of a living organism to construct out of mental elements and fragmentary experiences, concepts and notions of a higher order. The fuller elaboration of these statements is left for the succeeding paragraphs.

2. By "analysis" we mean the capacity for observing or discovering parts or differences in objects or qualities which for themselves seem unitary.

3. Although one can only speculate on these matters, it seems reasonable that analysis and synthesis are but the head and tail of a single function-tendency. This "analytic-synthetic" activity may be regarded a fundamental property characteristic of all irritable tissue, and more markedly so of nervous tissue. All forms and degrees of this function-tendency seem possible, from the simplest to the most complex.

2. PSYCHOLOGICAL ORIENTATION

If one thumb through the average psychology for a description and an analysis of the mental process of synthesization, integration, fusion, discrimination, differentiation, comparison, combination, blending, assimilation, and analysis, one's search will prove rather barren. Yet, as we shall have occasion to observe in greater detail later, these processes appear most fundamental for normal mental development.

The elements or primary structures of consciousness escape our observation. What we experience, what we

observe when we introspect, are fused or synthesized products of these elements. It has been said with justice that "Whoever would understand the facts of experience must know how they are built up out of the combination of elements."[1] Take mental functioning anywhere along the line, from the simplest type of perception to the most complex higher thought process, and at every point evidences of synthetic activity are observable or may safely be postulated. Percepts are synthesized sensations, ideas are synthesized percepts, images are recalled traces of these synthesized products, coördinations are orderly combinations of reflexes, judgments are orderly combinations of ideas, reasoning results from an orderly combination of judgments — emotion, will, learning, habit, all are combinations of some elementary form of mental process, these conscious states becoming more complex with the increase in complexity of the integrative pattern.

"Consciousness is not a shower of shot,"[2] a mass of disconnected elementary states, but at every stage it consists of a fusion of elementary processes into those more complex, and these finally into a functional unity. Psychology may well be defined as the science devoted to a study of mental elements, sensations, images, and affections, and the manner, condition, laws, concomitants, and effects of their combination at different developmental levels, phylogenetic and ontogenetic, normal and abnormal. This is a task alike for the structuralist,

[1] Judd, 13; 28. All references are to the bibliography at the end of Part A.

[2] Royce, 24; 108.

the behaviorist, the functionalist. It cannot be pre-empted by one, nor completely dealt with except by all.

At the bottom of this conception of mental elaboration is an implication which approaches the speculative. We cannot, however, on that account escape the force of its bearing on a reasonable explanation of observed conditions. In chemistry and physics the phenomenon of crystallization is thoroughly familiar. Given a fluid containing the proper ingredients, under certain conditions crystallization will take place. The canons of science justify our inquiry into the "how" but not the "why." Nevertheless, the postulation of a force or of forces, closely bound up perhaps with the atoms or the molecules, acting concertedly under fixed law and resulting in a new product, is a postulation which few will attempt to deny scientific scrutiny. So, too, in considering the mechanics of mental elaboration, to postulate a fundamental force or forces making for higher degrees of mental organization and for varieties of higher thought forms need not be considered irrelevant. The fundamental "ability" to "synthesize" or "combine" is not one which we may "will" to possess or not. We "synthesize" entirely independent of our personal choice in the matter. Just as we can exercise little if any control on the hyperfunction or hypofunction of some of the other organs of the body — for example, the kidney, the thyroid, the intestines — so also is our interference with brain function nominal. It seems evident that if one is born with good mental endowment his brain will "synthesize" to a degree and in a manner impossible to one whose native mental endowment is poor. Of course,

it is left, more or less, to individual initiative to exercise this inherited endowment efficiently. And yet, tests of "completion" or "combination" may measure this "synthesizing ability" in a very rough fashion, quite independent of schooling or other formalized training. Differences in level of mental ability may some day be explained, among other things, on the basis of differences in fundamental synthesizing ability, or the capacity of the nervous system to fuse elementary states of consciousness into higher thought forms. And perhaps the differences between a superior intelligence and a deficient intelligence may be found largely due to differences in this original basic synthesizing capacity.

Hobhouse has stated:

> "Where there is mind there is order and system, correlation and proportion, a harmonising of forces, an interconnection of parts. The organism which is gifted with intelligence shows it by arranging its actions on a certain plan. It adapts means to ends, which is one sort of correlation, and in so doing it perhaps brings a past experience to bear, interpreting a perception, for example, by memory; and this is another sort of correlation."[1]

Wherever we find traces of mental functioning there we inevitably observe the organization and correlation of mental elements into hierarchies. And this organization of mind on any functional level is evidenced by a parallel organization of behavior. In general, it might be stated that organized action is a product of intelligence, unorganized action the result of the absence or non-use of intelligence. And throughout the realm of mental func-

[1] Hobhouse, 10; 6.

tioning we note rather clearly the tendency for the fusion of elemental, conscious states into more complex forms, forms which serve more readily for individual adaptation to environmental demands.

"Our assimilations have not merely to do with the processes of perception and memory: they appear on the highest level of the intellectual life. *All our thinking involves assimilation.* When a novel object puzzles us, or when a problem baffles us, that is because we have not yet learned to assimilate the new experience to our former fashions of conduct. But when our puzzle is thoughtfully satisfied, this occurs because *we have learned to assimilate the new facts to the old principles*, i.e., to adjust our former methods of conduct, with a minimum of change, to the new situation. When the problem is solved, that is because what baffled us about a question which was asked, but to which we could not respond, disappears, because we have assimilated the matter at issue by remembering from our former experience an answer that serves the purpose. To be sure, such assimilation may be accompanied with alterations of habits that will need to be considered later under the head of Mental Initiative. *But every thoughtful process is, in at least one respect, a process of assimilation.*" [1]

In what manner is this concept of a synthesizing, or an assimilative, or combinative, or integrative principle different from our earlier and now discarded view that "associations of ideas" serve as a complete explanation of the varieties of mental experience? Our present hypothesis is rather one which assumes nervous tissue to possess the property or tendency of becoming integrated into higher units of complexity, as well as the tendency

[1] Royce, 24; 245.

for organizing environmental stimuli which appeal to various sense modalities into a consistent, understandable, and assimilable unity. The "association of ideas" is perhaps the most obvious superficial result of this synthetic-analytic tendency. It should be kept in mind, however, that "associations" evidence the *result* and not the *process* of this fundamental integrative force. The *process* is inherent in nervous tissue, and the *result* is present in consciousness as "associations between ideas." Titchener admirably summarizes the situation in seven words: "The brain associates, and meanings are associated."[1]

In general, it may be said that "the entire process of elaborating . . . objects of perception by the senses, presupposes for its explanation a constant activity of the mind . . . in combining the sensations into ever more complex forms. This combining activity is best called '*synthetic*,' or '*constructive*.'"[2] Nevertheless, there is an analyzing activity of mind which apparently is as important as the synthesizing. And perhaps analysis and synthesis are merely two aspects of nerve cell mechanism. Although antonyms, and expressing opposite tendencies, analysis and synthesis show evidences of each reënforcing and aiding the other.

Our aim is not to make hair-splitting distinctions between the processes of synthesis, assimilation, blending, combination, completion, induction, fusion, on the one hand, and analysis, comparison, deduction, differentiation, discrimination, on the other. It is sufficient to

[1] Titchener, 29; 149.

[2] Ladd and Woodworth, 14; 384.

state that the process or processes to which these terms refer are fundamental and may represent innate, dynamic potentialities characteristic of all nervous tissue. Many have attempted to differentiate the analytic activity from the synthetic. Our immediate purpose, however, will not necessitate any other assumption than that brain substance fundamentally manifests this activity — analysis and synthesis. Whether there are different "types"[1] is irrelevant to the present discussion. It perhaps is true, however, that at different age levels the character of the analyses and syntheses is different, the processes themselves changing as one passes through infancy to childhood, to adolescence, to adulthood, and finally through senescence.

Thorndike remarks that "the intellectual life of man seems to consist as much in discriminating, abstracting, taking apart, as in associating or connecting."[2] Although "fusion of some kind is always present in experience,"[3] nevertheless, the other aspect cannot be overlooked:

"Closely allied to the fact of focalness of thinking is the fact of analysis, the fact of breaking up a total fact into its elements, parts, or aspects. It is only as a result of such a process of breaking up total facts into their qualities that the elements of color, size, shape, weight, pressure, and the like are felt, in place of a 'big, blooming, buzzing confusion.' It is only as a result of such a process that many feelings of meanings and of intellectual relationships arise at all."[4]

[1] Meumann, 17.
[2] Thorndike, 27; 35
[3] Judd (1st ed.), 13; 137.
[4] Thorndike, 28; 105.

James clearly and with due emphasis maintains: "After discrimination, association!"[1] In speaking of these two processes, he says:

"It is obvious that the advance of our knowledge *must* consist of both operations; for objects at first appearing as wholes are analyzed into parts, and objects appearing separately are brought together and appear as new compound wholes to the mind. Analysis and synthesis are thus the incessantly alternating mental activities, a stroke of the one preparing the way for a stroke of the other, much as, in walking, a man's two legs are alternately brought into use, both being indispensable for any orderly advance." [2]

Hobhouse expresses a somewhat similar idea:

"Analysis rests on comparison, which is an act of synthesis, since it brings different experiences into relation; and gains in explicitness, *pari passu* with the common character on which the comparison turns. And not only does the concept rest on a synthesis, but its essential function is to make a further synthesis possible . . . There is thus in the free usage of detached concepts a synthetic process always at work articulating what has been disarticulated by analysis." [3]

The process of comparison stands out clearly in many writings and experiments as perhaps the most important feature of the analyzing procedure.

"By deliberate comparison I mean a mental confronting of the two objects, and a transition of attention from the one to the other, so as to discover some respect in which similar things differ in spite of their similarity, or in which different things agree in spite of their diversity,

[1] James, 11; 550. [2] James, 11; 550. [3] Hobhouse, 10; 294.

and also a fixing of the precise nature of this agreement or difference."[1]

Stout has so ably described this activity of "comparing" that this lengthy quotation may be pardoned:

"Comparison in all but a most rudimentary form is an ideational activity. Even when the objects compared are both present to the senses, each is scrutinised in turn. For anything more than a vague awareness of resemblance or difference, it is necessary to keep before the mind the ideal representation of the one object in the very act of examining the other. Only in this way can each detail and characteristic in turn be selected for comparison, so as to distinguish the points of difference from the points of agreement. Hence we may attribute the absence of comparison in animals in all but its most vague and rudimentary form, to the absence or extremely imperfect development of ideational activity in general.

"When the process of deliberate comparison plays an important part in the mental life, it involves a corresponding development in conceptual thinking, in the distinction of the general or universal from the particular. To compare is always to compare in some special respect. Some theoretical or practical end is to be subserved by the comparison. The difference or agreement to be discovered is not *any* difference or agreement, but one which has significance for the guidance of conduct or for the solution of a theoretical difficulty. Thus comparison takes place only in regard to the characteristics which happen to be interesting at the moment, other characteristics being disregarded or set aside as unimportant. Objects in other ways most diverse may yet in a certain respect be compared and found more or less similar, and objects in other ways most similar may be compared in a certain respect and found more or less unlike."[2]

[1] Stout, 26; 452. [2] Stout, 26; 455–456.

It may be of interest to recall at this point that Binet, with his unusually clear insight into the problems related to the development of intelligence, realized that the observation of the ways in which two things are alike is of as great a diagnostic value in determining intelligence level as the observation of ways in which two things are different. For that reason he devised tests for the purpose of determining a child's capacity to note similarities as well as differences. Terman, recognizing the importance of both processes, has included tests which measure both capacities in his revision and extension of the Binet scale. These tests, however, deal with concrete objects. Additional tests might be devised to measure the ability to note differences, using abstract material. Thus,

"To note the fall of an apple is to analyze out a comparatively obvious feature; to distinguish falling as a feature of the moon's revolution about the earth is still analysis, proceeding in much the same manner, but dealing with more difficult material and working with much more refined and elaborate tools." [1]

That there are individual differences in ability to compare and in ability to note similarities and differences seems obvious. According to James:

"Some people are far more sensitive to resemblances and far more ready to point out wherein they consist, than others are. They are the wits, the poets, the inventors, the scientific men, the practical geniuses. A native talent for perceiving analogies is reckoned by Professor Bain ('Study of Character,' p. 317) and by

[1] Ladd and Woodworth, 14; 606–607.

others before and after him, as the leading fact in genius of every order."[1]

The extent to which two things which are compared may appear different, may run from one extreme to the other. Two occurrences or things or facts may be so strikingly different that discrimination is relatively easy. On the other hand, the difference in the detail of the two objects under inspection may be so fine as to escape the notice of all but the most mature or the most expert.

"We all cease analyzing the world at some point, and notice no more differences. The last units with which we stop are our objective elements of being. Those of a dog are different from those of a Humboldt; those of a practical man from those of a metaphysician. . . . And when the omitted things are discovered and the unnoticed things are laid bare, it is not that the old thoughts split up, but that new thoughts supersede them, which make new judgments about the same objective world."[2]

A phenomenon closely linked with the matters mentioned above, and of considerable significance for mental progress, is that to which Mach has called attention,[3] namely, the inborn nature of experimentation in man — not the organized, thoroughly planned experimentation of the scientist, but the searching and the testing, the manipulating and the trying so characteristic of any normally developing child.

An item of great relevancy at this point is the question of individual differences in this experimenting attitude, the differences in the character of the observations (analyses), and the differences in the final syntheses.

[1] James, 11; 530. [2] James, 11; 489. [3] Mach, 15; 180.

Let us assume two children observing an object for the first time, one of superior and one of inferior mental ability. The little objective data which we now possess tend to show that the later imagery of this experience is markedly different in the two. The initial attitude is different. The observational procedure is different: the analyzing, the comparing, the selecting, the assimilating processes are different. "This selective activity not only differs from individual to individual, but even in the same individual it varies with the progress of his development and of his knowledge."[1] It is evident, therefore, that the final syntheses or products of observation will be different in the two. Many students of the graphic functions have frequently pointed out the development of a child's ability in drawing through distinct stages. Thus, the first phase is that of purposeless drawing — a mere scribbling for scribbling's sake; details, if observed at all, are detached entities, synthesis appearing only in its crudest form. Then follows a second stage, in which the child attempts a reproduction of objects seen, but synthetic incapacity is evident through inability to evaluate properly and place the numerous details into a reasonable whole. Thus, in drawing a picture of a human being the head is of enormous proportions, the limbs being appended from the chin, while other important parts of the body are omitted. In the final stage keenness of perception, a well-developed capacity for discrimination of detail, a knowledge and an experience of the laws of perspective, harmony, and balance, together with rigid self-criticism, are operative.

[1] Meumann, 16; 10, 11.

These steps in the development of mental ability are more or less duplicated in any variety of specific fields. For example, an average child of three can do no more than merely enumerate objects in a picture: the relationships between objects are not perceived. At seven, however, he can observe that the details bear some relation to one another and that some activity is portrayed. Only relationships of an obvious sort are observed. While at twelve, not only does he observe certain activities, but he obtains something more than is directly presented: he reads an interpretation into the picture, and perceives the subtle meanings underlying a given situation. Each higher stage requires a higher degree of mental synthesis, of combinative power. So, again, the same is true of ability to copy a circle, a square, a diamond, and a design. Each one of these figures seems to require a higher form of analytic-synthetic ability for successful performance. It is in this respect that the feeble-minded are characteristically defective. The consummation of associations or synthesizations of higher order, especially if of an abstract nature, are very difficult if not impossible for them.

Meumann summarizes the differences observable in children of different age levels in this capacity to elaborate and assimilate their worldly experiences.[1] He distinguishes three levels. On the first, synthesis is most predominant. Each object is cognized as a unit, its detail being distinguished separately in but a very inadequate fashion. A true analytic tendency is not yet observable. Whatever meager characteristics are noted

[1] Meumann, 17; 116 ff.

by the child are combined into a most illusory and fantastic notion. It is this stage which Meumann designates as the "Imaginative-Synthetic Period." This stage ends at about the age of eight or nine. Then follows a period which is characterized by a pronounced tendency toward analysis. The parts, peculiarities, and relationships of things are more attentively observed. The child develops a more truthful, a more accurate knowledge of objects. Imaginative combinations, elaborations, and interpretations become far less frequent. At about adolescence another period follows which is again synthetic, but of distinctly different character. The trend now is toward logical, reasonable combinations of impressions characteristic of the adult. "In that first period the child lives as in a fairy-world, in the third the sensing of reality appears normally active."[1]

There have been many attempts to divide "mental types" into the synthetic and the analytic. Whatever one's own reaction to the question of "types," it is nevertheless of interest to note some of the generalizations which have been made. Rignano[2] distinguishes between the synthetic and analytic "mentality" as follows: the possessor of the first is the discoverer of new concepts, he prefers comparison, analogies, the study of the broader relationships of cause and effect, and gives us an insight into the broad realm of any given field of endeavor. The analyst, on the other hand, is reflective, patient, persevering, deductive in his method; an experimenter, he prefers long, patient reasoning, and calculations complicated and almost unending. He appears to ad-

[1] Meumann, 17; 117. [2] Rignano, 22; 96.

vance prudently and surely, taking but a single step at a time. Among the synthesizers Rignano would place men such as Galileo, Newton, Faraday, Darwin, Comte, and among the analyzers, Watt, Stephenson, and Marconi.

Maudsley distinguishes between assimilative and discriminative types. The first observes small, delicate *resemblances* which are imperceptible to others, the latter possesses predominantly the capacity for recognizing points of *difference*.[1]

Again, Meumann, who has devoted so much of his discussion of individual differences to "types," distinguishes between the "analytic type" of person who is especially capable in the solution of single problems, in exact and analytic observation, a critic, a classifier — and the synthetic type, who is the builder of systems, one who manipulates and fashions whole regions of science.[2] The analyzer is concerned more with *differences*, the synthesizer more with similarities, with the search for similarity and analogy.[3] The first is "keen-sensed," the latter is "deep-sensed" or especially capable of noting wider and wider relationships. The highest type is that in which both attain their highest development in a single individual.

Meumann distinguishes still other types, which need not concern us for the present.

Some data from animal and genetic psychology have an application to the questions under discussion.

Biologists have almost entirely abandoned the search

[1] Maudsley, "The Physiology of Mind," 283.

[2] Meumann, 17; 159 ff.

[3] Meumann, 17; 161.

for that elusive if existent principle or trait which separates plants from animals. It would be rather generally conceded that all organisms manifest to some degree practically all the traits now claimed for higher animals.[1] The difference has been maintained to be one merely of degree or proportion. This view seems to be especially agreeable to those Weismannian disciples who believe that the primordial organism possessed, in the form of germinal determiners, all the potentialities for the development of all the higher species.

From the simplest one-celled organism up to man, Jennings finds evidence of perception, discrimination, choice, attention, fatigue, desire, pleasure, pain, fear, memory, habit, and even intelligence.[2] Of course, the lower down the scale the animal, the simpler are its mental mechanisms. But the germs of man's psychic activity are assumed to exist as far down the phylogenetic line as the protozoa. There is a string of unbroken continuity between the behavior of lower and higher species. Hobhouse distinguishes four stages in mental evolution, each representing a separate degree of mental correlation and adaptation, which basically reduces to differences in the character and development of the processes of synthesis and analysis at each of the four levels.[3] Ladd and Woodworth conclude that "greater power of analysis or discrimination belongs to man"[4] as compared with animals.

[1] Tashiro, "A Chemical Sign of Life."
[2] Jennings, 12.
[3] Hobhouse, 10.
[4] Ladd and Woodworth, 14; 555.

All experiments upon learning, animal and human, have clearly indicated the importance of attention, discrimination, and association in this process. This is illustrated in the attempt of Ladd and Woodworth to term the "trial and error" learning in animals "learning by varied reaction through selection of the successful variants."[1] Without variation in behavior-habit the right method would not have been hit upon by the animal in his learning efforts. And without attention, discrimination, and association, even if chance success favored the animal, nothing would be gained which would be of future service in the same situation.

James cites an instance where a dog, arriving with its master before a rowboat which they were in the habit of using, found it full of water. A sponge was generally used to bale this out. On this occasion the sponge had been left at home, about a third of a mile away. The master, disliking to make the trip, made gestures of cleaning out the boat, at the same time saying to the dog, "Sponge, sponge; go get the sponge." The dog ran off and brought back the desired article. To James this was evidence of nothing but contiguous association. If the dog had been unable to find the sponge and had brought back a dipper or a mop instead, "such a substitution would have shown that embedded in the very different appearances of these articles, he had been able to discriminate the identical partial attribute of capacity to take up water, and had reflected, 'For the present purpose they are identical.' "[2] It is this power to analyze

[1] Ladd and Woodworth, 14; 550.

[2] James, 11; 349–350.

out a detail common to a series of objects, acts, or phenomena, which possibly differentiates man from brute, and the superior among men in intelligence from those who are inferior. This process has its analogue in abstraction. Miss Fisher found that the experience of similarity "was the most important component detail in the process of generalizing abstraction"[1] and that the "experience of similarity and that of generality were in themselves strikingly akin."[2]

In the case of animals it is a question to what extent varieties of past experience are interpreted in the light of a new situation. Thorndike doubts whether animals possess such generalized elements of thought as ideas. And although Hobhouse's attitude is strongly affirmative, nevertheless

"if we attribute ideas to an animal, they are not ideas arrived at by any breaking-up, analysis, or other elaboration of what is given in perception. None of my animals (with the possible exception now and again of the monkeys) showed the least understanding of the how or why of their actions, as distinct from the crude fact that to do such and such a thing produced the result they required. It is this want of what one may call analysis, that made, for example, the push-back bolt such a difficulty. What Jack or the elephant knew was, crudely, that they had to push this bolt. That the reason why they had to push it was to get it clear of the staple they obviously never grasped."[3]

How remarkably analogous is the mental functioning of the animal as here pictured with that of the high-grade moron, who knows what is right and what wrong, but

[1] Fisher, 7; 97. [2] Fisher, 7; 208. [3] Hobhouse, 10; 200–201

fails to grasp the "why." Is it a question of synthetic-analytic incapacity? We shall have occasion to return to this matter again later.

"If then, positively, these experiments suggest 'practical ideas,' negatively, they strongly suggest absence of any sort of analysis in the genesis of those ideas. An animal can shift its attention to this or that object, or change within the sphere of perception; but it cannot apparently follow out the structure of any complex object with any minuteness and accuracy."[1]

Some contributions from the field of genetic psychology are illuminating. We are frequently told that the child deals with the concrete; the adult with the abstract. This means, presumably, that the child deals with the isolated, with objects as they have been immediately perceived — the adult, however, has generalized and has abstracted common properties and deals with these idealizations by means of word symbols or physiological set. To experience the meaning of "twenty," or of "promise," or of "opposite," or of "justice," is apparently beyond the power of children's mental capacity. Progress in this respect comes with chronological age and with the character of native mental endowment. It is quite apparent that "In mental growth connection and analysis, association and dissociation, putting things together and breaking things up into parts, constantly work together."[2]

There are a few pedagogical implications worth mentioning with regard to this synthetic-analytic activity. Education, according to Royce, involves, first, "acts of

[1] Hobhouse, 10; 201 [2] Thorndike, 28; 219–220

sensory observation, of recalling images, of repeating words, of drawing diagrams, of performing experiments, and so on, indefinitely. Then we acquire gradually the power to 'survey at a glance,'" and "this process of surveying at a glance involves a high degree of differentiation of our simultaneous conscious states."[1] This ability to bring to consciousness a mass of related facts, data, information, experiences, to pick out the essentials necessary for the purposes of the particular moment, or to select and epitomize, or summarize relevant details for interpretation as a unit — that is truly mental functioning close to its apex of development. Meumann emphasizes the importance of combinative ability when he speaks of difficulties in work of the schoolroom, in geography, for example:

> "Defective capacity to learn geography may be due either to sub-normal ability to deal with concrete visual imagery (maps and the like); or when normal visual imagery and normal memory of names is present, it may be due to an inability to combine concrete visual images with the auditory-motor images of words."[2]

In the aphasias we possess excellent illustrations of total absence of specific combinative potentialities, which of course have clearly demarked neurological bases. Royce has conceived the educative process as "an attainment of synthesis by means of analysis."[3] And analysis "is manifestly one of the most incessantly performed of all our mental processes."[4]

By emphasizing the importance of analysis and synthe-

[1] Royce, 24; 254.
[2] Meumann, 16; 204.
[3] Royce, 24; 258
[4] James, 11; 502.

sis for normal mental functioning we are not claiming a berth for a new "faculty." Evidence of every kind is again and again demonstrating the unity of mind, and the integrative character of its physical substrate: the nervous system. We cannot localize "analysis" and "synthesis." These tendencies seem to be properties inherent in nervous protoplasm, present the very moment an aggregate of nerve cells become organized into any form of functional unit. This fact emphasizes again Wundt's declaration that

> "there is, in reality, but one psychical center; and that is the brain as a whole, with all its organs. For in any at all complicated psychical process, these organs are brought into action, if not all together, at any rate over so wide a range and in such various quarters as to forbid the delimitation of special psychical centers within the functional whole."[1]

And one wonders if James was not correct when, in discussing "analysis" he apologized for the short space devoted to this activity and insisted: "I think I emphasize it enough when I call it one of the ultimate foundation-pillars of the intellectual life."[2]

3. NEUROLOGICAL ORIENTATION

Our discussion of mental mechanisms invariably forces us to a consideration of the anatomical and the physiological structures underlying them. Comparative psychology has often emphasized the slight differences in the sensory capacities between man and animal. And wherever the difference has been emphasized it has

[1] Wundt, 34; 218. [2] James, 11; 530.

often been stressed in favor of animals. Our studies of the feeble-minded reveal slight, if any differences, in the sensory capacities of normal and mentally deficient. It seems that the factors which make for intelligent behavior do not reside in a refined sense-organ, necessarily, although refined sense-organs are extremely desirable. Intelligence is something cerebral. We, therefore, naturally look for the clearest exemplification of the activity of the analytic-synthetic tendency in the function of the cerebrum.

Most anatomists and neurologists agree that no true nervous elements are evident in our progress from lower to higher organisms until we reach the coelenterates. The Protozoa possess no nervous system. The Metazoa do. Jennings finds no evidence to explain the behavior of the Metazoa as different on the basis of their possession of a nervous system. What the Metazoa do, the Protozoa can also do. "The possession of a nervous system brings with it no observable essential changes in the nature of behavior. We have found no important additional features in the behavior when the nervous system is added."[1] In Vorticella, Jennings found that the receptive end is ciliated peristome. Stimuli of certain kinds will be transmitted and contractions of the myoid filament at the fixed end of the cell will be noticeable. In multicellular organisms the same simplicity of nervous mechanism is observed, with this exception: increased complexity of the nervous fabric is correlated with greater delicacy of individual readjustment to changes in environmental conditions. What is of espe-

[1] Jennings, 12; 263.

cial significance is the fact that "the peculiar properties of the nerve cells are properties of protoplasm in general, but somewhat accentuated."[1] Each cell of our body, although there are comparatively few like itself, nevertheless is capable of uniting in function with others strikingly different in structure. The nature of those conditions making for unity of function out of a diversity of structure, certainly implies a synthesis of a most subtle nature, true for somatic protoplasm as well as for nervous.

The work of Sherrington, epoch-making in its contributions toward our understanding of nervous mechanisms, throws some light on our question.

"The nervous system in its simplest forms is diffuse — a number of scattered mechanisms performing merely local operations with much autonomy save that they have communication with their immediate neighbors across near boundaries.[2] . . . It is ill suited, therefore, to produce the integration of a large and complex individual as a whole.[3] . . . Yet the coördination it brings about in its own local field may be strikingly effective."[4] The earthworm is a good example of a segmented or metameric nervous system. Each metamere leads an almost independent existence. But "the integrative function of the nervous system is seen to perfection in the welding together of metameres into the unity of an animal individual."[5] There are hierarchies of nervous integration paralleled by hierarchies of functional organization.

"By longitudinal integration short series of adjoining

[1] Jennings, 12; 280.
[2] Sherrington, 25; 311.
[3] Sherrington, 25; 311.
[4] Sherrington, 25; 312.
[5] Sherrington, 25; 314.

segments become in respect to some one character combined together, so as to form in respect to that character practically a single organ.[1] . . . And it is in the integration of long series, or of the whole series, of segments one with another, that, apart from psychical phenomena, the nervous system seems to reach its acme of achievement."[2]

Jennings has presented a great mass of evidence demonstrating the constant "feeling about" of all lower organisms, the attempting of new directions, the retiring from those which are harmful, and the persisting in those which are of advantage to the organism.[3] This feeling about, this avoiding and persisting, force one to grant some degree of discrimination to the animal, and once that is granted, the existence of a fundamental analytic-synthetic element logically follows.

Present-day evidence has demonstrated that the neuron is an anatomic, genetic, physiologic, trophic, and functional unit. Each neuron is a complete cell in and of itself. Nevertheless, the complexity of interconnection between neurons is beyond human comprehension. From the simplest reflex to the most complex higher thought process a basic tendency of "get-together-ness" and "stay-together-ness" is observable among neurons. And where it is absent or defective, mental progress from infancy to adulthood can hardly be conceived.

The passage of stimuli through sensory pathways and the complexity of interconnection between what Sher-

[1] Sherrington, 25; 344.

[2] Sherrington, 25; 344.

[3] See Jennings' description of the behavior of an amoeba in pursuing another for the purpose of devouring it. — Jennings, 12; 16, 17.

rington designates the "silent areas," illustrate the complexity underlying mental mechanism. The terminal nuclei and the thalamus are among the important way-stations to cerebral connection. But

"the thalamus, since it possesses many short fibres connecting its own parts, is probably something more than a mere way-station. Apparently, the sensory impulses from the different receptors come together here and join in such a way that the impulses which pass from here to the cortex are already organized to a certain extent."[1]

What function other "relay" stations or connecting links play in synthesizing and organizing primary mental states can only be conjectured. Throughout his book Sherrington stresses the interconnecting or "integrative" function of the nervous system, which makes for a unity of action out of a diversity of organs.[2]

In running through the evidence presented on the mental changes which result from cortical ablations, one is amazed by the contradictory results which have been found in various quarters. Careful experimentation in the last few years, notably that of Franz in this country, is demonstrating with greater and greater conclusiveness that the integrity of the cortex is necessary for the integrity of mental functioning. Meynert makes a good deal of the fact that disorganization of any portion of the cortex manifests itself in some form of psychic disturbance.[3] Morat stresses the fact that the brain may

[1] Ladd and Woodworth, 14; 93.
[2] Sherrington, 25.
[3] "Populäre Vorträge," 2 ff.

be regarded "a prodigy of unity" as well as "a prodigy of complexity."[1] Mach concludes that "A review of anatomic, physiologic, and psychopathologic findings forces one to the conclusion, that the integrity of consciousness is dependent upon the integrity of the cortex."[2] And finally, the careful work of Franz leads him to the generalization that, "Both the clinical and the physiological evidence points to a dependence of certain mental processes on the integrity of the frontal lobes and, in fact, on the integrity of the brain as a whole."[3]

4. SUMMARY

To summarize: Whether we investigate the purely mental aspects of our behavior, or whether we observe the integrating of nerve element with nerve element, we are struck with the significance of synthesis for normal mental functioning. We shall have occasion in a later chapter to consider the current definitions of intelligence, and with this new orientation we shall suggest another definition which perhaps will take us back to something more fundamental than "adaptation."

[1] Morat, 18; 1. [2] Mach, 15; 42. [3] Franz, 8; 64.

Bibliography

1. Bartlett, F. C., "An Experimental Study of Some Problems of Perceiving and Imaging;" *British Journal of Psychology*, 1916, 8: 222–267.
2. Bergson, Henri, "Creative Evolution." (New York; Holt & Co., 1913; 407 pp.; transl. by A. Mitchell.)
3. Bolton, J. S., "The Brain in Health and Disease." (London, Arnold, 1914; 479 pp.)
4. Book, W. F., "Psychology of Skill." (University of Montana Publication in Psychology.)
5. Bryan, W. L., and Harter, N., "Studies in Physiology and Psychology of the Telegraphic Languages," *Psychological Review*, 1899, 6: 345–375.
6. Duprat, L., "Association mentale et causalité psychologique." *Revue Philosophique*, 1913, 75: 452–470.
7. Fisher, Sara Carolyn, "The Process of Generalizing Abstraction; and Its Product, The General Concept." (*Psychological Monographs*, 1916, vol. 21, No. 2; 213 pp.)
8. Franz, Shepherd Ivory, "On the Functions of the Cerebrum; the Frontal Lobes." (*Archives of Psychology*, 1907, vol. 1, No. 2; 64 pp.)
9. Gregor, A., "Untersuchungen über die Entwicklung einfacher logischer Leistungen (Begriffserklärung). (*Zsch. f. angew. Psychol.*, 1915, 10: 339–451.)
10. Hobhouse, L. T., "Mind in Evolution." (London, Macmillan, 1901; 415 pp.)
11. James, W., "The Principles of Psychology." (New York, Holt & Co., 1890; vol. 1, 689 pp.; vol. 2, 704 pp.)
12. Jennings, H. S., "Behavior of the Lower Organisms." (New York, Columbia University Press, 1915; 366 pp.)

13. Judd, Charles H., "Psychology; General Introduction." (Boston, Ginn & Co., 1917, [2d. ed.]; 358 pp.)
14. Ladd, George T., and Woodworth, Robert S., "Elements of Physiological Psychology." (New York, *Scribner's*, 1911; 704 pp.)
15. Mach, E., "Erkenntnis und Irrtum." (Leipzig, Barth, 1905; 461 pp.)
16. Meumann, E., "The Psychology of Learning." (New York, D. Appleton & Co., 1913; 393 pp.; transl. by J. W. Baird.)
17. Meumann, E., "Vorlesungen zur Einführung in die experimentelle Pädagogik." (Leipzig, Engelmann, 1911; 725 pp.)
18. Morat, J. P., "Physiology of the Nervous System." (London, Constable & Co., 1914; 680 pp.)
19. Pillsbury, W. B., "Attention." (London, Swan Sonnenschein & Co., 1908; 346 pp.)
20. Ribot, Th., "Essay on the Creative Imagination." (Chicago, Open Court Publishing Co., 1906; 370 pp.; transl. by A. H. N. Baron.)
21. Ribot, Th., "Evolution of General Ideas." (Chicago, Open Court Publishing Co., 1899; 231 pp.; transl. by F. A. Welby.)
22. Rignano, Eugenio, "Les diverses mentalités logiques," *Scientia*, 1917, 22: 95–125.
23. Rignano, Eugenio: Qu'est-ce que le raisonnement? *Scientia*, 1913, 13 (supplement): 30–57.
24. Royce, Josiah, "Outlines of Psychology." (New York, Macmillan, 1904; 392 pp.)
25. Sherrington, C. S., "The Integrative Action of the Nervous System." (New York, Scribner's, 1906; 411 pp.)
26. Stout, G. F., "A Manual of Psychology." (New York, Hinds, Noble & Eldridge, 1899; 643 pp.)

27. Thorndike, E. L., "Educational Psychology." Vol. 2: "The Psychology of Learning." (New York, Teachers College, Columbia University, 1913; 452 pp.)
28. Thorndike, E. L., "The Elements of Psychology." (New York, Seiler, 1905; 351 pp.)
29. Titchener, E. B., "A Beginner's Psychology." (New York, Macmillan, 1916; 362 pp.)
30. Titchener, E. B., "Experimental Psychology of the Thought-Processes." (New York, Macmillan, 1909; 318 pp.)
31. Watson, John B., "Behavior." (New York, Holt, 1914; 439 pp.)
32. Wundt, W., "Grundzüge der physiologischen Psychologie." (1910, 3 vols. [6 ed.])
33. Wundt, W., "Lectures on Human and Animal Psychology." (454 pp.)
34. Wundt, W., "Principles of Physiological Psychology." (London, Swan Sonnenschein & Co., 1904; 347 pp.)
35. Ziehen, Th., "Leitfaden der physiologischen Psychologie." (Jena, Gustav Fischer, 1914 (Zehnte Auflage); 504 pp.)

PART B

THE BLOCK-DESIGN TEST

THE BLOCK-DESIGN TEST[1]

CHAPTER I

Statistics Regarding the Subjects who Were Given the Block-design Tests

In the following division we shall present tables giving the chronological ages of the subjects, arranged by source; the Binet mental ages of the subjects, arranged by source, the Binet mental ages and the intelligence quotients arranged by sex and source; the correspondence of chronological and mental ages, totaled separately for sex and source; the age-grade distribution, classified by sex; an age-progress table, classified by sex; nationality; place of birth; social class; home conditions; and teachers' estimates of intelligence. The questionnaire on p. 61 was utilized to obtain most of this information. The aim was to make certain that the group selected was typical, and if atypical, in what direction and to what extent. The results seem to show that the children selected were representative of the larger distribution of children at large and consequently the final norms seem valid. Mayfield yielded children somewhat inferior, Palo Alto yielded children somewhat superior, and Menlo Park more of an average type.

Table I shows the life ages of the children tested. At almost every year the cases number between 20 and 30, and the years range from 4 to 16 in the case of the public school children, and from 9 to about 37 in the case

[1] The complete set of material may be obtained from C. H. Stoelting Co., 3037–3047 Carroll Ave., Chicago, Illinois.

TABLE I

	Chronological Age																Boys	Girls	Total
	4–7 to 5–6	5–7 to 6–6	6–7 to 7–6	7–7 to 8–6	8–7 to 9–6	9–7 to 10–6	10–7 to 11–6	11–7 to 12–6	12–7 to 13–6	13–7 to 14–6	14–7 to 15–6	15–7 to 16–6	16–7 to 17–6	17–7 to 22–6	22–7 to 37–6	37–7 to over			
Mayfield	1	2 5	5 4	2 5	2 3	8 8	5 3	3 5	3 9	4 4	4 4	2	2				41	52	93
Palo Alto High								1	2 1	4 2	4 3	5 6	2				18	12	30
P.A. Intermed.							1	4 5	3 1	3 1	1 1	1					12	9	21
P. A. Lytton		1 2	6 2	4 7	6 5	7 4	4 3	3 2	1 1	1							33	26	59
Menlo Park		3 2	3 7	4 6	8 3	4 2	8 6	4 2	7 5	6 3	1 1	2	2				50	39	89
	1	6 9	14 13	10 18	16 11	19 14	17 13	15 14	16 17	18 10	10 9	8 8	4 2				154	138	292
Normal Total P. S. Group	1	15	27	28	27	33	30	29	33	28	19	16	6						
Vineland					1		2	2	2	2 1		1 1	2 1	5 3	3 1	1	14	14	28
Eldridge						1			2	2	1	1	2 2	5 11	8 9	2 1	24	23	47
					1	1	2	2	2 2	4 1	1	2 1	4 3	10 14	11 10	2 2	38	37	75
F. M. Total					1	1	2	2	4	5	1	3	7	24	21	4			
	1	6 9	14 13	10 18	17 11	20 14	17 15	15 16	18 19	22 11	11 9	10 9	8 5	10 14	11 10	2 2			
Grand Total	1	15	27	28	28	34	32	31	37	33	20	19	13	24	21	4	192	175	367

Male . . . Light figures   Female . . . Dark figures

of the feeble-minded. The Mayfield cases numbered 93, the Palo Alto 110, the Menlo Park 89. There are 28 feeble-minded cases included from the Vineland, N. J., Training School, and 47 feeble-minded cases from the California State Home for the Feeble-minded, making a grand total of 367 cases. Each of these was given the block designs, the two Trabue tests (B and C), the Binet test, and the dissected sentences test. The time spent in the examination of each case varied from one hour to three hours, the average being about two hours. Almost all of the examinations required two sittings, in some cases three. At no single time was a subject kept over an hour and a half. Conditions for examination were always excellent. The writer, who did all the examining, always had a room apart, free from disturbances and distractions.

In Table II are given the mental ages of the subjects examined. It will be observed that most of the ages are clustered between 6 and 13. All considered, this is an excellent range, the number of cases at each age varying from 20 to 50.

In Table III are given the Mental Ages and I. Q.'s for the public school boys and girls. The modal I. Q. for both groups, separately and combined, is 100. Table IV gives the I. Q.'s for the feeble-minded. The median I. Q. for the feeble-minded falls in the 50 I. Q. group.

It is evident from these tables that there is a positive correlation between increase of mental age and increase of intelligence quotient. Supposedly we would expect *no* correlation. But our school group is evidently selected. In the grammar grades there has been some

TABLE II

	Mental Age																	Boys	Girls	Total
	2–7 to 3–6	3–7 to 4–6	4–7 to 5–6	5–7 to 6–6	6–7 to 7–6	7–7 to 8–6	8–7 to 9–6	9–7 to 10–6	10–7 to 11–6	11–7 to 12–6	12–7 to 13–6	13–7 to 14–6	14–7 to 15–6	15–7 to 16–6	16–7 to 17–6	17–7 to 18–6	18–7 to 19–6			
Mayfield		1	2	1 5	5 7	9 8	4 4	4 11	3 6	7 3	3 3	1 2	1 2	1				41	52	93
P. A. High												2	2 3	5 2	4 3	3 1	4 1	18	12	30
P. A. Inter.									3 3	3 3	2	4 2		1				12	9	21
P. A. Lytt.			1	2 2	4 1	4 6	5 6	7 3	3 6	2	3 1	2	1					33	26	59
Menlo Park				3 6	3 4	7 8	7 3	10 3	8 5	6 4	4 3	1 2	1	1				50	39	89
		1	3	6 13	12 12	20 22	16 13	21 17	17 20	18 10	12 7	8 8	4 6	5 5	4 3	3 1	4 1	154	138	292
P. S. Group		1	3	19	24	42	29	38	37	28	19	16	10	10	7	4	5			
Vineland	1 2	3 2	1 1	2 3	1 2	1 1	2 2	2	1	1								14	14	28
Eldridge		1	5 4	3 4	1 1	3 4	3 1	2 2	4 2	2	2	1	1	1				24	23	47
	1 2	4 2	6 5	5 7	2 3	4 5	5 3	4 2	4 3	1 2	2	1	1	1				38	37	75
F. M. Total	3	6	11	12	5	9	8	6	7	3	2	1	1	1						
	1 2	5 2	9 5	11 20	14 15	24 27	21 16	25 19	21 23	19 12	12 9	8 9	5 6	6 5	4 3	3 1	4 1			
Grand Total	3	7	14	31	29	51	37	44	44	31	21*	17	11	11	7	4	5	192	175	367

Male . . . Light figures Female . . . Dark figures

* One case no block-designs given.

TABLE III — P. S. GROUP
MENTAL AGE AND I. Q.

MENTAL AGE	INTELLIGENCE QUOTIENTS 26–35		36–45		46–55		56–65		66–75		76–85		86–95		96–105		106–115		116–125		126–135		136–145		TOTAL	
2–7 to 3–6																									0	0
3–7 to 4–6			1																						1	0
4–7 to 5–6											2				1										3	0
5–7 to 6–6										1		2	4	5	2	5									6	13
6–7 to 7–6							1	1	2	2		2	6	3	3	4									12	12
7–7 to 8–6							1	1			4	3	5	3	4	6	5	4	1	5					20	22
8–7 to 9–6									2	1	4	2	2	3	7	5	1	2							16	13
9–7 to 10–6								1	1	2	4	4	7	3	4	5	3	1	2	1					21	17
10–7 to 11–6									2	1	4	5	3	7	5	4	3	2		1					17	20
11–7 to 12–6										1	5	2	6	2	3	4		1	3						18	10
12–7 to 13–6									1		3	2	2	1	3	3	4			1					12	7
13–7 to 14–6														2	4	1	1	4	2	1	1				8	8
14–7 to 15–6													1	1	2	2	1	2				1			4	6
15–7 to 16–6															3	1		3	2	1					5	5
16–7 to 17–6																1	2	1	1			1	1		4	3
17–7 to 18–6																	1		1	1	1				3	1
18–7 to 19–6																				1	2		2		4	1
			1				2	3	8	8	26	22	36	30	41	41	21	20	12	12	4	2	3		154	138
Total			1				5		16		48		66		82		41		24		6		3		292	
															98–91											

Male . . . Light figures Female . . . Dark figures
MODAL GROUP = I. Q. 100

elimination of the mentally retarded, and this is especially striking in the high school grades. Our schools seem to demand higher mental ages and higher I. Q.'s for progress through the grades. In the primary grades I. Q.'s fall as low as 56. In the intermediate school there was only one case below 76, and in the high school there was

TABLE IV

F. M. GROUP

	Intelligence Quotients									
	20	30	40	50	60	70	80	90	100	
Vineland . . .	1	8	5	7	5	1	1			
Eldridge . . .		9	8	9	7	9	2	2	1	
Total . . .	1	17	13	16	12	10	3	2	1	75

MEDIAN I. Q. = 50

not a single case with an I. Q. less than 86. Another item of importance: the upper right-hand section of Table III includes children who are accelerated mentally but retarded pedagogically, and the lower left-hand section of the table includes children who are retarded mentally, but accelerated pedagogically.

In the public school group, 189 out of the 292 have I. Q.'s between 86 and 115 — in other words, 64.7 per cent have average intelligence.

Table IV gives the distribution of I. Q.'s for the feeble-minded cases. And Table V summarizes the totals in Tables III and IV.

Tables VI to XII present in detail the I. Q. distributions for each source, with its median I. Q. The median I. Q.'s (crude) for the various schools were,

TABLE V

MENTAL AGE AND I. Q.

P. S. AND F. M. GROUPS COMBINED

(Boys and Girls)

	16 TO 25	26 TO 35	36 TO 45	46 TO 55	56 TO 65	66 TO 75	76 TO 85	86 TO 95	96 TO 105	106 TO 115	116 TO 125	126 TO 135	136 TO 145	TOTAL
P. S. Group			1		5	16	48	66	82	41	24	6	3	292
F. M. Group	1	17	13	16	12	10	3	2	1					75
Total	1	17	14	16	17	26	51	68	83	41	24	6	3	367
								192						

Mayfield, 90; Palo Alto High, 110; Palo Alto Intermediate, 100; Palo Alto Lytton, 100; Menlo Park, 90; Vineland, 45; Eldridge, 50.

TABLE VI

MAYFIELD CASES

Mental Age	Intelligence Quotients																				Total	
	40		50		60		70		80		90		100		110		120					
4	1																				1	
5									1				1								2	
6								1			1	1		3							1	5
7					1		2	2		2	1	1	1	2							5	7
8					1				2	2	3	1		3	3	1		1			9	8
9							1	1	2	1		2			1						4	4
10						1		2	1	3	2		1	4		1					4	11
11							1		1	3		2	1	1							3	6
12									3	2	1		1	1			2				7	3
13									1	2	2			1							3	3
14													1			2					1	2
15														1	1	1					1	2
16																1						1
	1				2	1	4	6	11	15	10	7	6	16	5	6	2	1			41	52
Total	1				3		10		26		17		22		11		3				93	

MEDIAN I. Q. = 90

TABLE VII

PALO ALTO INTERMEDIATE

Mental Age	Intelligence Quotients												Total	
	70		80		90		100		110		120			
11 . .	1		1	1		2	1						3	3
12 . .			1		2	1		2					3	3
13 . .							1		1				2	
14 . .							1		1	1	2	1	4	2
15 . .														
16 . .										1				1
	1		2	1	2	3	3	2	2	2	2	1	12	9
Total .	1		3		5		5		4		3		21	

MEDIAN I. Q. = 100

Boys . . . Light figures Girls . . . Dark figures

TABLE VIII

Palo Alto High

Mental Age	Intelligence Quotients												Total	
	90		100		110		120		130		140			
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
14 . .		2												2
15 . .	1	1	1	1		1							2	3
16 . .			3			1	2	1					5	2
17 . .				1	2	1	1			1	1		4	3
18 . .					1		1	1	1				3	1
19 . .								1	2		2		4	1
	1	3	4	2	3	3	4	3	3	1	3		18	12
Total .	4		6		6		7		4		3		30	

Median I. Q. = 110

Boys . . . Light figures Girls . . . Dark figures

TABLE IX

Palo Alto Lytton

Mental Age	Intelligence Quotients												Total	
	80		90		100		110		120		130			
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
5 . .	1												1	
6 . .		1	1		1	1							2	2
7 . .			2		2	1							4	1
8 . .		1	1	1	2	1		1	1	2			4	6
9 . .		1	1	1	4	3		1					5	6
10 . .		1	2		3	1	2			1			7	3
11 . .			1	1	1	2	1	2		1			3	6
12 . .					1				1				2	
13 . .							3			1			3	1
14 . .					1						1		2	
15 . .												1		1
	1	4	8	3	15	9	6	4	2	5	1	1	33	26
Total .	5		11		24		10		7		2		59	

Median I. Q. = 100

Boys . . . Light figures Girls . . . Dark figures

TABLE X

Menlo Park

Mental Age	Intelligence Quotients														Total	
	60		70		80		90		100		110		120			
6 . . .						1	2	4	1	1					3	6
7 . . .		1					3	2		1					3	4
8 . . .		1			2		1	1	2	2	2	2		2	7	8
9 . . .			1		2		1		3	2		1			7	3
10 . . .			1		3		3	3			1		2		10	3
11 . . .				1	2	1	2	2	2	1	2				8	5
12 . . .			1	1	1		3	1	1	1		1			6	4
13 . . .					2			1	2	2					4	3
14 . . .									1	1		1			1	2
15 . . .									1						1	
16 . . .										1						1
		2	3	2	12	2	15	14	13	12	5	5	2	2	50	39
Total .	2		5		14		29		25		10		4		89	

Median I. Q. = 90

TABLE XI

Eldridge

Mental Age	Intelligence Quotients																Total	
	30		40		50		60		70		80		90		100			
4 . . .	1																1	
5 . . .	4	4	1														5	4
6 . . .			1	4	1		1										3	4
7 . . .			1	1													1	1
8 . . .					3	4											3	4
9 . . .						1	2		1								3	1
10 . . .							2	2									2	2
11 . . .									4	2							4	2
12 . . .										2								2
13 . . .												2						2
14 . . .														1				1
15 . . .													1				1	
16 . . .															1		1	
	5	4	3	5	4	5	5	2	5	4		2	1	1	1		24	23
Total . .	9		8		9		7		9		2		2		1		47	

Median I. Q. = 50

Boys . . . Light figures Girls . . . Dark figures

TABLE XII

VINELAND

Mental Age	Intelligence Quotients 20		30		40		50		60		70		80		Total	
3 . . .	1			2											1	2
4 . . .			2	2			1								3	2
5 . . .			1	1											1	1
6 . . .					2	2		1							2	3
7 . . .						1	1			1					1	2
8 . . .							1			1					1	1
9 . . .							2	1		1					2	2
10 . . .									2						2	
11 . . .												1				1
12 . . .													1		1	
	1		3	5	2	3	5	2	2	3		1	1		14	14
Total .	1		8		5		7		5		1		1		28	

MEDIAN I. Q. = 45

Boys . . . Light figures Girls . . . Dark figures

In Table XIII are presented the mental age and chronological age distributions for the public school group. It is evident from this table that there is a somewhat larger percentage of those who are retarded mentally than accelerated: 36 (12 per cent) are accelerated more than one year, and 81 (28 per cent) are retarded more than one year. In Table XIV the data of Table XIII are summarized for years of retardation and acceleration. It will be observed that 59 per cent are at age mentally, or within one year; 22 per cent are within two years; 9 per cent are within three years; 7 per cent are within four years; 2 per cent within five; and ½ of 1 per cent within six years of their mental age.

TABLE XIII

CHRONOLOGICAL AND MENTAL AGES

(P. S. GROUP)

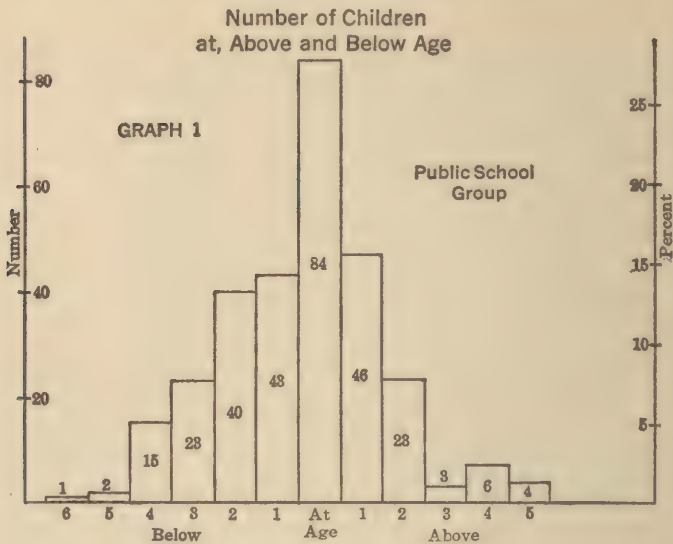
CHRONOLOGICAL AGE	MENTAL AGE																																				TOTAL
	3		4		5		6		7		8		9		10		11		12		13		14		15		16		17		18		19				
5					1																														1		1
6					1		5	7		1		1																							6	9	15
7					1		1	3	5	5	7	5																							14	13	27
8			1					3	4	1	2	9	2	3	1	2																			10	18	28
9									1	2	4	2	6	5	4		1	2																	16	11	27
10									1	1	6	4	2	2	5	4	2	2	2			1	1												19	14	33
11									1	2			3		6	4	4	4	2	2	1					1									17	13	30
12											1		1	2	2	1	2	4	3	3	3	1	2	3					1						15	14	29
13												1	2	1	3	4	2	4	3	1	3	[illegible]	1	2		1	1			1	1				16	17	33
14																1	4	3	3	1	1		4	1	2	2	1	2	1				2		18	10	28
15																	1	1	2	2	3	2			1	1	1	1		1			2	1	10	9	19
16																	1		2	1				2	1	1	1	2	1	1	2	1			8	8	16
17																1			1		1	1					1		1						4	2	6
			1		3		6	13	12	12	20	22	16	13	21	17	17	20	18	10	12	7	8	8	4	6	5	5	4	3	3	1	4	1			
Total			1		3		19		24		42		29		38		37		28		19		16		10		10		7		4		5		154	138	292

Boys . . . Light figures Girls . . . Dark figures

TABLE XIV

Mental Age		Boys	Girls	Total	
Above	5 yrs.	4		4	
	4 "	3	3	6	
	3 "	2	1	3	36
	2 "	10	13	23	
	1 "	26	20	46	
At Age		41	43	84	173
Below	1 yr.	25	18	43	
	2 yrs.	21	19	40	
	3 "	12	13	25	
	4 "	9	6	15	83
	5 "	1	1	2	
	6 "		1	1	
Total		154	138	292	

In Graph 1 is presented graphically the retardation and acceleration of the 292 public school cases. A skew is observable toward the right. This is again manifest in Graph 2, which graphically represents the distribution of I. Q.'s for the same group of children. In making a comparison between the grouped distributions of I. Q. obtained by Terman in his revision of the Binet for 905 cases and our results, we note the following: Terman found 35 per cent having I. Q.'s above average, in our cases 25.4 per cent were above average; he also found 33.9 per cent average (96 — 105 I. Q.), in our cases 28 per cent were average; and he found 31.3 per cent below average, while in our cases 46.6 per cent were below average. Our inferior group is 15 per cent larger, the average group is about 6 per cent smaller, and the superior group is about 10 per cent smaller. The fact that Terman excluded the foreign-born, which have been included in our data and amount to about 5 per


Number of Children
at, Above and Below Age
GRAPH 1
Public School
Group
Number
Percent
80
60
40
20
25
20
15
10
5
1
2
15
23
40
43
84
46
23
3
6
4
6
5
4
3
2
1
At
Age
1
2
3
4
5
Below
Above

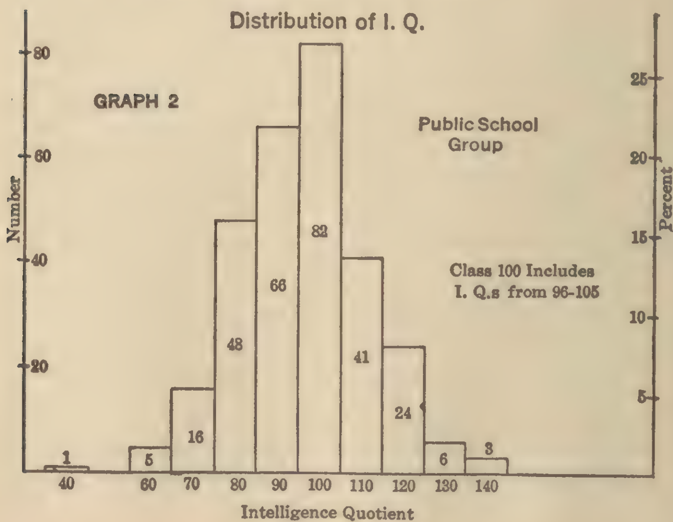

Distribution of I. Q.
GRAPH 2
Public School
Group
Class 100 Includes
I. Q.s from 96-105
Number
Percent
80
60
40
20
25
20
15
10
5
1
5
16
48
66
82
41
24
6
3
40
60
70
80
90
100
110
120
130
140
Intelligence Quotient

TABLE XV

AGE-GRADE DISTRIBUTION

Chronological Age	Grade																						Total	
	1		2		3		4		5		6		7		8		I		II		III			
5	1																						1	
6	6	9																					6	
7	9	7	5	5		1																	14	13
8		3	5	8	1	7																	10	18
9	1		2	2	8	7	3	2	2														16	11
10			1	1	8	3	5	3	3	6	1	1											19	14
11					2	1	5	2	8	5	2	3		2									17	13
12					1		2	1	4	1	3	4	3	7	1	1	1						15	14
13							2		2	2	4	3	3	9	3	1	2	2					16	17
14									1		3	1	3	4	7	3	4	2					18	10
15													2	1		5	2	1	1	2	1		10	9
16															3	2	2	3	3	2		1	8	8
17															2	2			2				4	2
	21	19	4	16	20	19	17	8	20	14	13	12	11	23	20	14	11	8	6	4	1	1	1 4	138
Total	40		30		39		25		34		25		34		34		19		10		2		292	

Boys . . . Light figures Girls . . . Dark figures

cent of the public school cases, would help account for the discrepancy. We have also made no eliminations because of race or nationality. As Pintner has so ably stated, the mere fact of large numbers of cases is no safeguard for reliability of standardization; and in obtaining accurate norms, individuals of all classes and degrees should be included.[1] We have stressed careful technique rather than multiplicity of cases, or perfection in sampling.

Table XV is the age-grade distribution of our public school cases. What is striking here is the practically symmetrical distribution as compared with the skewed curve for the mental ages and I. Q.'s. This is clearly brought out in Table XVI, which summarizes briefly the data of Table XV. It will be observed that 35 per cent are at age pedagogically, 31 per cent are accelerated, and 34 per cent are retarded. The percentage at or within one year of grade is 83 per cent.

TABLE XVI

		Boys	Girls	Total	Per Cent
Above grade	3 yrs.	1		1	.4
	2 yrs.	8	6	14	4.7
	1 yr.	31	45	76	26.0
At age for grade		53	49	102	34.9
Below grade	1 yr.	38	26	64	21.9
	2 yrs.	17	9	26	9.0
	3 yrs.	6	2	8	2.7
	4 yrs.				
	5 yrs.		1	1	.4
	Total	154	138	292	100.0

[1] See R. Pintner and D. G. Paterson, "A Scale of Performance Tests," pp. 73 ff.

TABLE XVII

GRADE-PROGRESS DISTRIBUTION

Grades	Years in School													Total
	Less than One	1	2	3	4	5	6	7	8	9	10	11	12	
1 . .	11 8	3 8	7 3											21 19
2 . .		1 2	12 12	1 2		1								14 17
3 . .			8	15 9	5 2									20 19
4 . .				7 2	6 4	4 2								17 8
5 . .				2	8 6	5 2	0 3	1	1					20 14
6 . .					1	5 8	4 2	2 1	1 1					13 12
7 . .						2	6 7	3 11	1 1	1 2				11 23
8 . .							2 2	8 6	6 4	3 1	1	1		20 14
I .						2		3	7 1	2 3				11 7
II .								1	1 1	1	4 2			6 4
III .									1 1					1 1
	11 8	4 10	19 23	23 15	20 12	16 15	18 14	13 23	18 9	7 6	4 3	1		154 138
Total .	19	14	42	38	32	31	32	36	27	13	7	1		292

Boys . . . Light figures Girls . . . Dark figures

Tables XVII and XVIII are concerned with the correlations between grade and progress, or the number of years spent in school: 40 per cent had made normal progress, 41.5 per cent had made accelerated, and 18.5 per cent had made retarded progress. The percentage

TABLE XVIII

		Boys	Girls	Total	Per Cent
Accelerated Progress	4 yrs.	2		2	.7
	3 yrs.	1	2	3	1.1
	2 yrs.	4	10	14	4.7
	1 yr.	54	48	102	34.9
Normal progress		60	57	117	40.0
Retarded Progress	1 yr.	29	15	44	15.1
	2 yrs.	2	5	7	2.4
	3 yrs.	2	1	3	1.1
	Total	154	138	292	100.0

making normal progress, or within one year of it, amounted to 90 per cent. The numbers making rapid progress are unusually large. This is accounted for by the fact that a good part of the testing came in February and March (1918) when children had just been promoted to a new grade. They were thus given credit for progress in that grade in which they had not yet made any progress at all. Taking this into account, the grade-progress distribution is of about the same type as the age-grade.

In Table XIX are given the numbers and per cents of boys and girls who are at grade or within one year of it by chronological age (termed "in the right grade"), and who are between 86 to 115 I. Q. (termed "right I. Q."). It is clear from the table that 60 per cent of the

TABLE XIX

GRADE LOCATION AND I. Q.

	Number		Per Cents		Total	Per Cent
	Boys	Girls	Boys	Girls		
In right grade and of right I. Q.. .	91	85	59%	62%	176	60%
Neither right grade nor right I. Q. .	25	12	16%	9%	37	13%
Right I. Q. only .	7	6	5%	4%	13	4%
Right grade only .	31	35	20%	25%	66	23%
Total . . .	154	138	100%	100%	292	100%
	$\underbrace{\qquad\qquad}_{292}$					

By "Right Grade" is meant at grade or within one year of it for chronological age.

By "Right I. Q." is meant an I. Q. between 86 to 115.

children are what may be termed "average" and these help stabilize any tendency toward deviation from the absolute norm.

In Table XX is given a list of geographical divisions and the number claiming descent from the nations of that region. Ranked for frequency, the leading nationalities came in the following order: North European, American, and combinations of North European and American. These accounted for over 80 per cent of all the cases. The descendants from northern groups numbered 80.4 per cent, from southern groups 11.7 per cent, Orientals 3.4 per cent and from those not stating 4.5 per cent.

The data on birthplace by states, summarized for geographical division, and for birthplace in California

or outside of the state, are here presented: 74 per cent were born in California, 26 per cent were born elsewhere. The Pacific Coast States contributed 76.5 per cent, Mountain States 5.2 per cent, East North Central 4.5

TABLE XX

NATIONALITY AND BIRTHPLACE

	Number	Per Cent
(1) North European	108	37.0
(2) South European	22	7.6
(3) American	91	31.1
(4) Oriental	10	3.4
(5) American and N. European	36	12.3
(6) American and S. European	5	1.7
(7) N. European and S. European	7	2.4
(8) ? or not stated	13	4.5
Total	292	100.0

per cent. The smallest numbers came from the East South Central States (.7 per cent) and from the New England States (0 per cent).

Tables XXI to XXIII and Graphs 3 to 5 give the results of teachers' estimates of intelligence, home conditions, and social class. The writer has devised a series of tables for grouped measures of variability which are given in the appendix and which are here utilized for the purpose of comparing the percentages actually found with what one would expect theoretically. It will be noted that the average discrepancy for intelligence estimation is only 2 per cent, for estimation of home conditions 3 per cent, and for social status 4 per cent. This increase is to be expected, for to teachers, intellectual performance is under their observation

continually, and this is not as true for home conditions or social status.

TABLE XXI

TEACHERS' ESTIMATES

INTELLIGENCE

	Number	Per Cent	Theoretical Expectation	Discrepancy
Very inferior	4	1.4	0.2*	+1.2
Inferior	11	3.7	4.0	−0.3
Below average	52	17.8	24.0	−6.2
Average	133	45.5	43.6	+1.9
Above average	68	23.2	24.0	−0.8
Superior	21	7.3	4.0	+3.3
Very superior	3	1.1	0.2	+0.9
Total	292	100.0	100.0	

AVERAGE (A. M.) DISCREPANCY = 2.1%

* See Appendix I.

TABLE XXII

TEACHERS' ESTIMATES

HOME CONDITIONS

	Number	Per Cent	Theoretical Expectation	Discrepancy
Very inferior	2	0.7	0.2*	+0.5
Inferior	17	6.0	4.0	+2.0
Below average	57	20.2	24.0	−3.8
Average	113	40.0	43.6	−3.6
Above average	56	19.6	24.0	−4.4
Superior	26	9.2	4.0	+5.2
Very superior	12	4.3	0.2	+4.1
?	9			
Total	283	100.00	100.0	

AVERAGE DISCREPANCY = 3.4%

* See Appendix I.

TABLE XXIII

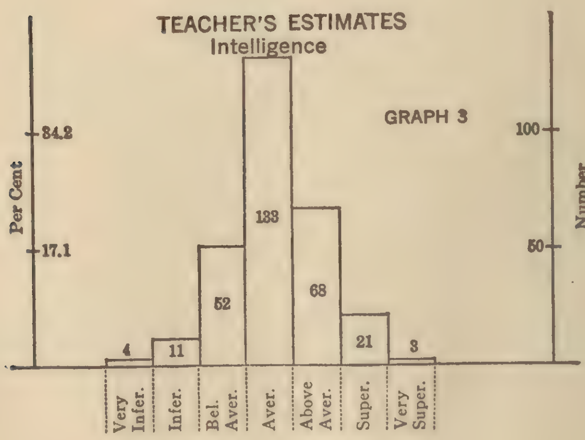
TEACHERS' ESTIMATES

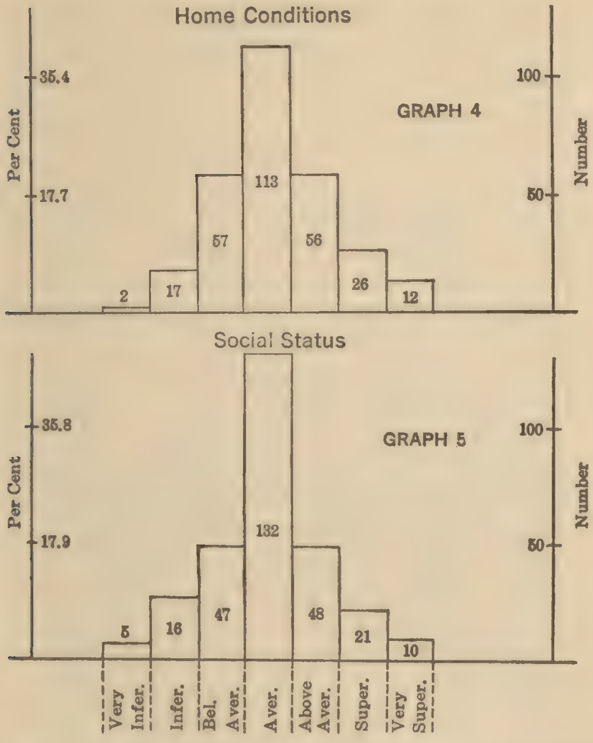
SOCIAL STATUS

	NUMBER	PER CENT	THEORETICAL EXPECTATION	DISCREPANCY
Very inferior	5	1.7	0.2*	+1.5
Inferior	16	5.7	4.0	+1.7
Below average	47	16.8	24.0	−7.2
Average	132	47.3	43.6	+3.7
Above average	48	17.4	24.0	−6.6
Superior	21	7.5	4.0	+3.5
Very superior	10	3.6	0.2	+3.4
?	13			
Total	279	100.0	100.0	

AVERAGE DISCREPANCY = 3.9%

* See Appendix I.





Summing it all up, we have tested a fairly representative group of school children, on the whole perhaps not markedly different from average children in other parts of this country.

Sample of Questionnaire

To the Teacher

The information rendered on this blank will be held strictly confidential, will be used for statistical purposes

only and without the mention of any child's name. The information will aid greatly in the standardization of a series of block-design tests.

1. Name of school..................................
2. Name of pupil..................................
3. Date of birth..................................
4. Grade (when test was made)......................
5. Years attended school...Grades skipped...Repeated...
6. Place of birth..................................
(what state if American, what country if foreign).
7. Descendant of what nationality......................
(E.g., Scotch-Irish, Portuguese, German-American, etc.)
8. Language spoken at home..........................
9. Occupation of father..............................
10. Quality of the child's work (as compared with average children for this grade). (Write appropriate number, see rating scale)..............................
11. Teacher's estimate of the child's intelligence (as compared with average children of the same age). (Write appropriate number, see rating scale)...............
12. Home conditions (consider neatness, size, comfort, provision of life necessaries, relations between the parents, parental supervision). (Write number)........

Rating Scale

In your ratings indicate

1. Very Inferior.......by 4
2. Inferior............by 5
3. Below Average......by 6
4. Average............by 7
5. Above Average.....by 8
6. Superior...........by 9
7. Very Superior......by 10

13. Social class (consider culture, intelligence, social position of the group from which the child comes).
(Write number)..................................

14. What is the character of the child's work in the following school subjects (use rating scale).
(Write number)..................................

1. Reading..............	8. Drawing.............
2. Writing...............	9. Nature Study........
3. Spelling...............	10. Music...............
4. Language............	11. Manual Training......
5. Arithmetic...........	12. Household Arts.......
6. History...............	13. Science..............
7. Geography...........	14. Deportment..........

15. If this child were given a mental test involving *chiefly* the *use of language*, or the ability to understand and handle language symbols (as contrasted with a test requiring silent observation and *performance*), would you expect him (her) to pass.....................

Very low
Low
Below Average
Average
Above Average
High
Very High

Note. — Put an X *before* the appropriate word

16. Additional information which will throw light on the child's intelligence, school success or failure, social status, peculiarities of any sort, etc. will be especially welcome and appreciated. *We thank you.*

CHAPTER II

The Test Material

I. THE MATERIAL

(*a*) *The Blocks*

The blocks which are used are manufactured by The Embossing Co., are called "Color Cubes," and may be secured at any of the large department stores and at the various distributing centers of Milton Bradley's. There are sixteen cubes of one inch dimension and all are painted as follows:

One side red,
One side blue,
One side white,
One side yellow,
One side blue and yellow (divided on the diagonal),
One side red and white (divided diagonally).

The character of the colors is indicated on the pages of designs printed in this monograph. A slight difficulty, experienced by possibly one or two subjects out of every one hundred, was a just perceptible but nevertheless disconcerting difference in shade between the blue and yellow on the full faces and the same colors on the diagonal sides. This can be remedied in the later standardization of the test material. One set of six blocks will last through the examination of from four to five hundred

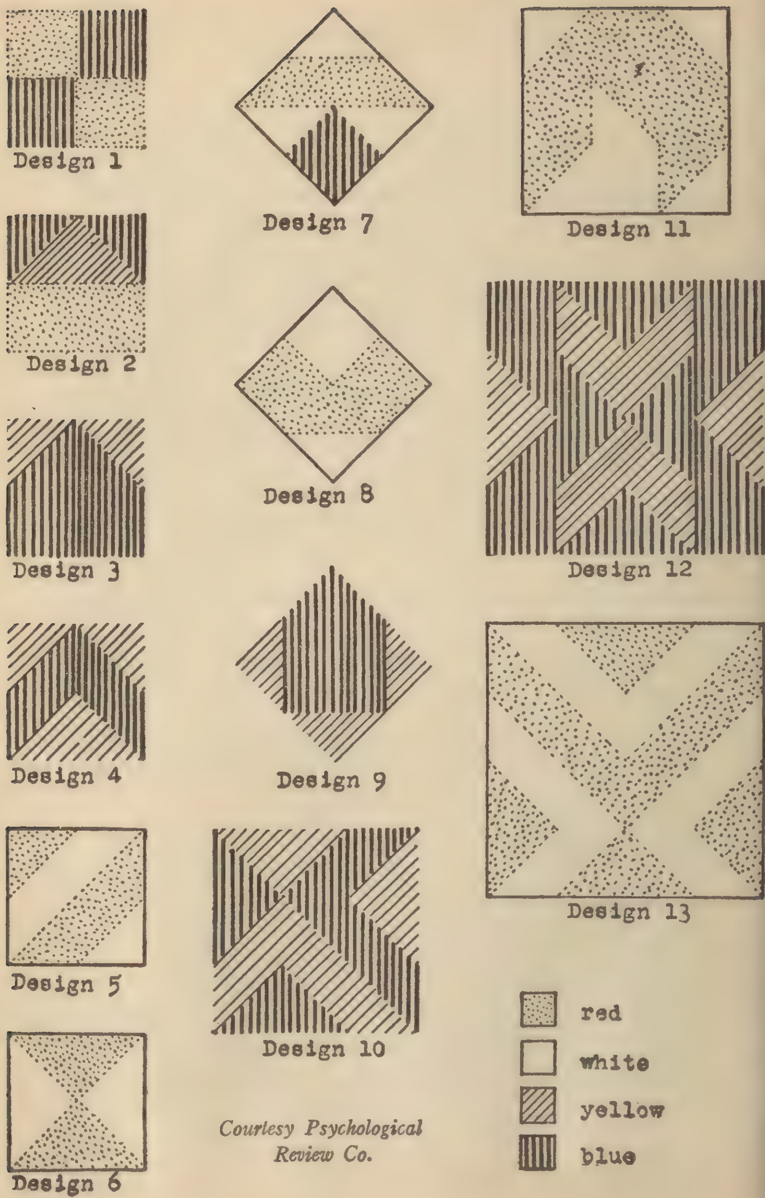
children without showing much wear and tear. After that the cubes can be repainted without difficulty.

It is interesting to watch the response of children and even adults when they are given colored cubes to handle. There is no doubt that an appeal exists which touches the roots of some very fundamental original tendencies. Of all the subjects tested, not one has manifested any absence of a desire to *combine* these cubes in some fashion. The experimenter needs only to direct this natural interest toward a specific end and then apply a scientific measuring technique to evaluate the results.

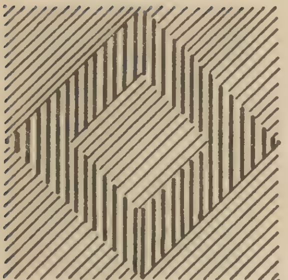
(*b*) *The Designs*

On pages 66–67 are presented the designs utilized in this study. The Roman numerals refer to the numbers given the designs during the time when the tests were being standardized. The Arabic numerals designate the final numbering of each design. The original number was thirty-five, but fifteen were eliminated in a few of the early preliminary testings. The designs are graded in difficulty which increases by modifying the designs at various stages in the following manner:

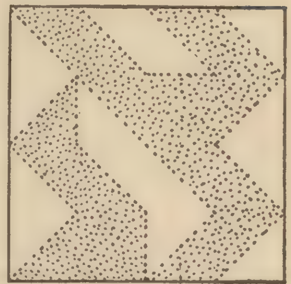
1. By the use of the full colors;
2. By the use of few diagonaled sides;
3. By the use of all diagonaled sides;
4. By turning the design on one of its corners;
5. By eliminating the outside boundary line;
6. By increasing the number of blocks to be used;
7. By increasing dissymmetry in design;
8. By decreasing the number of different colors used in each design.



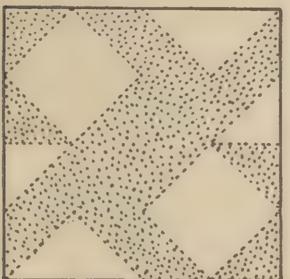
Courtesy Psychological Review Co.



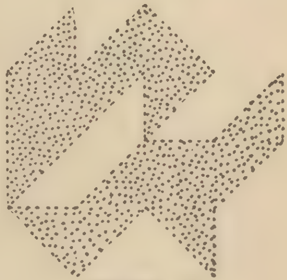
Design 14



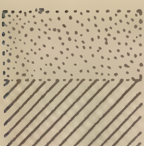
Design 16



Design 15

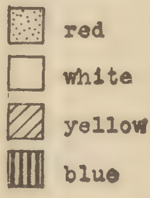


Design 17



Trial Design A

Courtesy Psychological Review Co.



To perform the test, utilizing the twenty designs, one averaged about an hour or an hour and a half. In the final revision three designs have been eliminated, leaving seventeen, thus decreasing somewhat the time necessary to apply the tests with no significant decrease of reliability. The criteria for rejection were based on correlations with those arrays of evidence presuming to yield an index of intelligence, such as is obtained through the use of the Binet scale, and also upon the basis of the diagnostic value of each design, determined by the progress of its curve with increasing chronological age. The results at present indicate that the block designs are as good as any single test in the Binet scale (though better, perhaps, in the sense of *diagnostic value*), as good as the Trabue Language Completion Tests, or any other similar single type test, whether involving the use of language or whether mere performance.

The designs, appropriately colored, are printed on medium-thick, white, semi-gloss cardboard. The dimensions of the card are three by four inches. The printed designs, placed in the center of the card, are *one-fourth* the size of the actual designs when the cubes are used. In other words, the face of a cube represented on the designs is only *one-half* of an inch on each of its sides. Thus design No. 1 is one inch square, design No. 10 is one and a half inches square, and design No. 14 is two inches square.

The writer has found it of assistance to place in the lower right-hand corner the time limit for each design. These values follow:

TABLE XXIV

TIME LIMITS FOR EACH DESIGN

DESIGN (NUMBER)	TIME LIMIT (MINUTES)	DESIGN (NUMBER)	TIME LIMIT (MINUTES)	DESIGN (NUMBER)	TIME LIMIT (MINUTES)
1 . .	1½	7 . .	2	13 . .	3½
2 . .	1½	8 . .	2	14 . .	3½
3 . .	1½	9 . .	2	15 . .	4
4 . .	2	10 . .	3	16 . .	4
5 . .	2	11 . .	3½	17 . .	4
6 . .	2	12 . .	3½		

The time limit set for each design is about one minute longer than the time within which a correct response may reasonably be expected.

It may be of interest to remark that if the full limit is allowed on each test the working-time totals only forty-five minutes for all of the seventeen designs. With practice an examination should average about thirty to forty minutes. In some cases it may take only fifteen or twenty minutes, in others perhaps an hour.

(c) *The Score Card and the Method of Scoring*

In the succeeding table are presented the score values of each of the seventeen designs and the number of score points to be deducted if a design is successfully completed with excess time and with excess moves.

To clarify the table, one or two illustrations will be utilized. For example, design No. 2 has a score value of 5. This full amount is attained if a reagent completes the design successfully in less than 31 seconds and with less than 7 moves. If 31 or more seconds are utilized, one point is deducted from the score, and if 7 or more

moves are made an additional point is deducted. Take again design No. 13, which has a score value of 9. This full amount is attained if the subject completes the design successfully in less than 2 minutes and 21 seconds, and with less than 31 moves. If completed between

TABLE XXV

Score Card

Design No.	Score Value Points	Points to be Subtracted		
		Excess Time		Excess Moves 1 Point
		1 Point	2 Points	
1	3	21″ and over		6 and over
2	5	31″ and over		7 and over
3	6	21″ to 35″	36″ and over	8 and over
4	7	31″ to 1′ 0″	1′ 1″ and over	10 and over
5	7	36″ to 1′ 5″	1′ 6″ and over	11 and over
6	7	36″ to 1′ 0″	1′ 1″ and over	12 and over
7	7	41″ to 1′10″	1′11″ and over	11 and over
8	8	41″ to 55″	56″ and over	10 and over
9	9	56″ to 1′10″	1′11″ and over	15 and over
10	9	1′56″ to 2′10″	2′11″ and over	22 and over
11	8	1′46″ to 2′30″	2′31″ and over	19 and over
12	9	2′26″ to 2′40″	2′41″ and over	30 and over
13	9	2′21″ to 2′33″	2′34″ and over	31 and over
14	9	2′26″ to 2′40″	2′41″ and over	32 and over
15	9	2′41″ to 3′ 0″	3′ 1″ and over	32 and over
16	10	2′41″ to 3′ 5″	3′ 6″ and over	31 and over
17	11	2′41″ to 2′55″	2′56″ and over	30 and over

Maximum score — 133 points.

2 minutes 21 seconds and 2 minutes 33 seconds, one point is deducted; and if 2 minutes 34 seconds or more are spent on the problem, two points are deducted. And if 31 or more moves are made an additional point is deducted from the score value of the design.

The scoring of a performance is a very simple matter. This will be self-evident from the following examples:

Example one: Design No. 7 successfully completed in 1 minute and 23 seconds and at the end of 9 moves. Score 7, for successful completion, less 2 points for excess time. Final score 5.

Example two: Design No. 10 successfully completed in 1 minute 48 seconds, and after 19 moves. Score 9, for successful completion. No deductions for excess time or excess moves. Final score, 9.

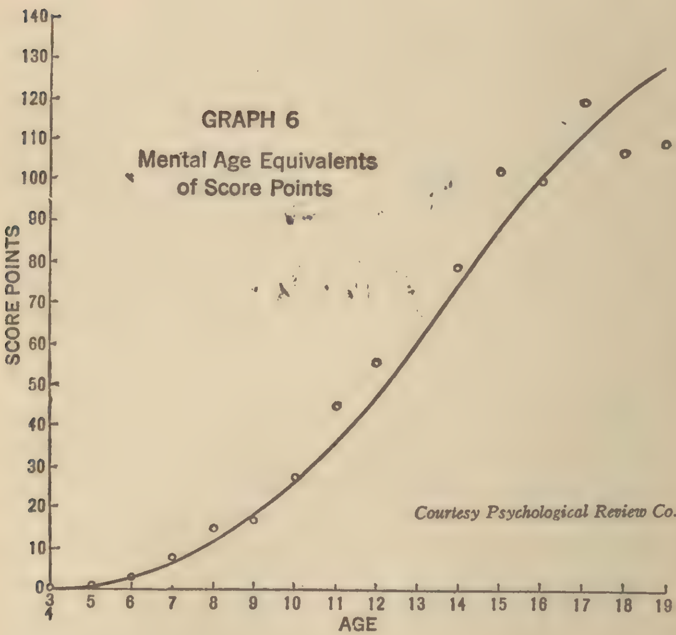
Example three: Design No. 16, successfully completed in 3 minutes 27 seconds, and after 48 moves. Score 10, for successful completion. Deduct 2 points for excess time, and one point for excess moves. Final score 7.

It may be worth remarking that successful performance, speed, and what may be termed accuracy are combined in the final score. Successful performance receives greatest weight, speed next, and accuracy next. The weight ratio as explained elsewhere in the monograph is roughly 4:2:1. This ratio has been empirically determined. The prevalent opinion that speed and accuracy cannot be combined in one score, does not hold with the Block-design Tests. The writer felt that success, speed, and accuracy each had its own diagnostic importance, and in order to make the tests most effective all should, and must, be taken into account in the final score summation. For a more detailed discussion of this matter see a later section.

(*d*) *The Norms*

The procedure involved in obtaining norms for the different designs was quite a complicated one, requiring a great deal of careful statistical work. In this effort

the writer utilized the currently accepted standardization methods, with but slight modification. An explanation of the general procedure utilized, together with a description of various methods of checking the results has



been left for a later chapter. Suffice it to say that the score points mentioned in Table XXV are to be interpreted in the same light as those of Buckingham in his standardization of his Spelling Tests, of Trabue in his standardization of his Language-Completion Tests, and

of Woody in his standardization of his Arithmetic Tests. In this section the final results, merely, will be presented.

Graph 6 is the curve indicating the scores to be expected at the various ages from 3 years to 19 years and 6 months. This curve has been smoothed but slightly

TABLE XXVI

MENTAL AGE EQUIVALENTS OF SCORE VALUES

Score Points	Mental Age	Score Points	Mental Age	Score Points	Mental Age	Score Points	Mental Age
0	5-3 or below	33	10-9	66	13-5	99	15-9
1	5-7	34	10-10	67	13-6	100	15-10
2	6-0	35	10-11	68	13-6	101	15-11
3	6-3	36	11-0	69	13-7	102	16-9
4	6-6	37	11-1	70	13-8	103	16-1
5	6-9	38	11-2	71	13-9	104	16-2
6	7-0	39	11-3	72	13-9	105	16-3
7	7-3	40	11-4	73	13-10	106	16-4
8	7-6	41	11-5	74	13-11	107	16-5
9	7-8	42	11-6	75	14-0	108	16-7
10	7-10	43	11-7	76	14-1	109	16-8
11	8-0	44	11-8	77	14-1	110	16-9
12	8-2	45	11-9	78	14-2	111	16-10
13	8-4	46	11-10	79	14-3	112	16-11
14	8-5	47	11-11	80	14-4	113	17-1
15	8-7	48	12-0	81	14-5	114	17-2
16	8-9	49	12-1	82	14-6	115	17-4
17	8-10	50	12-2	83	14-7	116	17-5
18	9-0	51	12-3	84	14-7	117	17-6
19	9-1	52	12-4	85	14-8	118	17-8
20	9-3	53	12-5	86	14-9	119	17-9
21	9-4	54	12-6	87	14-10	120	17-10
22	9-6	55	12-7	88	14-11	121	18-0
23	9-8	56	12-8	89	15-c	122	18-2
24	9-9	57	12-9	90	15-0	123	18-3
25	9-11	58	12-10	91	15-1	124	18-5
26	10-1	59	12-10	92	15-2	125	18-7
27	10-2	60	12-11	93	15-3	126	18-9
28	10-3	61	13-0	94	15-4	127	18-11
29	10-4	62	13-1	95	15-5	128	19-1
30	10-5	63	13-2	96	15-6	129	19-3
31	10-7	64	13-3	97	15-7	130	19-7
32	10-8	65	13-4	98	15-8	131	19-11

within the range of ages below 10, though rather considerably from 15 to 19. This was necessarily the result of a deficiency in the number of cases at the higher ages. The median score is represented by a circlet placed above each age.

In Table XXVI are presented the mental age equivalents of each score from 1 (mental age 5 years and 7 months) to 131 (mental age 19 years and 11 months). These values were derived from Graph 6.

2. THE DIRECTIONS FOR APPLYING THE TESTS

Preliminaries: Seat the S comfortably at a table, noting that his visual angle when working with the tests is not less than 45 degrees. Be sure that no designs are visible in your preliminary instructions, nor more than a single design at any one time. The blocks which are not being utilized should be kept in a box, apart, so that they are either invisible to the S, or if visible, the blocks should be arranged so that the top sides are all of the same color.

Method: Part 1. *For Subjects Who Can Understand Spoken Language*

(Section A) Take a block. (Instructions to S are placed in quotation marks. For design 1 four blocks will have been removed from the box.) "Here are some blocks — give me the name of the color on this side." Sides with the full colors are presented first. Place your finger on the side designated. After S has responded, turn to another side. "And what is the color on this side?" — "Now the color here?" — "And what is the

color here?" — If S has succeeded in naming the colors correctly, proceed with the experiment. (If he has failed, further instructions are given below, in Part 2.) Then the experimenter explains: "Now on this side we have blue and yellow (point), and on this side red and white (point). And all the blocks are painted in the same way."

(Section B) "What you are to do is this: Take these blocks" (shuffle them so that when finally placed before S, no more than one quarter of the blocks have topside colors which are present in the design, the separate blocks being placed apart, flat on table, and not piled one on top of another), "pick out the right colors, put them together, and make them look, on top, just like this" (point to design 1). Give no further hints or suggestions if the directions have been understood. *Caution:* Be sure that all the blocks are thoroughly shuffled *before* the design is presented. The purpose is to eliminate the possibility of studying the design before being ready to begin work with the cubes.

(Section C) If S has not understood what is meant, E may perform trial design (A)[1] slowly, using pantomime freely, S watching closely, after which S is requested to repeat the operation. This may be repeated any number of times until S understands. When he does, proceed with the designs in order, beginning at (Section B), and continuing with (Section D).

[1] Trial design (A) is represented on the pages with the other designs and is used only when under the provisions of Section C further preliminary explanation is necessary. Trial design A is a four-block design, two full white sides above, two full yellow sides below.

(Section D) After the first design has been completed or failed, the blocks are again shuffled, observing the cautions in (Section B), and the S is told again to "take these blocks, pick out the right colors, put them together, and make them look on top just like this." (Point to design No. 2.) These instructions remain the same for all the designs. The S is not told at any time the number of the blocks he is to use.

Record: Both time and moves are recorded. A move is counted when a block is given its initial position on the table. Each separate and distinct change in the position of a block is counted a move. Sometimes a child will make three or four changes in the position of a cube, the topside remaining the same color (especially true of diagonal sides, e.g. red — white). But each change in position is counted a separate move. If success is not attained within the time limit, no credit is assigned. The time limits are indicated on the design cards.

The whole test is not regarded as complete unless there are, ordinarily, at least five consecutive failures on designs after the last success, and where doubt exists as to the inability in the later designs, give as many designs beyond the last success as is deemed wise.

Part 2. *For Subjects Who Do Not Know the Names of the Colors*

Take all the blocks out of the box and place on the table so that the single-colored faces are all on the top side of the cubes. Have an equal number of reds, yellows, blues, and whites. Point to a red-topped

block and ask the child to point to all the blocks that have the same color on top. Do the same for the other three colors. If the child can distinguish the colors, proceed with the test at Section B.

Part 3. *For Subjects Who Cannot Understand Spoken Language*

By means of gestures and pantomime go through the procedure in Part 2. If S can distinguish the colors, proceed with Section C, and through the various designs. The method of recording remains the same.

CHAPTER III

Standardization Technique

1. The Year-Scale Method
2. The Percentile Method
3. The Point-Scale Method
4. The P. E. or Linear Projection Method

The past few years have seen such a rapid development in the technique of scale-making for the purpose of measuring mental capacities and pedagogical progress that any reasonable discussion of this branch of psychological and educational statistics would be entirely too extensive and consequently out of place in this monograph. However, there are four distinct methods which have had more or less ready application and it is to these alone that the writer will confine his discussion.

1. THE YEAR-SCALE METHOD

This is the method which has been utilized in standardizing the various versions of Binet-Simon Measuring Scale for Intelligence. After tests have been selected and applied to large numbers of unselected children of different ages, each test in the scale is considered separately and the percentage passing the test at the various life ages is determined. The fundamental assumption is that if a test is a measure of intelligence development, there will be an increase in the per cent passing the test with increase of age. (Increases due to differences of

maturity or experience apart from intelligence, are, of course, ruled out.) A criterion which Binet used is to locate a test at a specific age if about 75 per cent of the children at that age pass the test. Usually the percentage fluctuates between 60 and 85 or 90 per cent. There are a number of distinct objections to this arbitrary procedure, which have been pointed out by A. S. Otis [1] and T. L. Kelley.[2] A test, however, is generally regarded as most useful which shows the most rapid advance with age, and which is placed in the year level at which about 75 per cent begin to pass. Although the writer recognizes the superiority of the 50 per cent criterion, nevertheless to make the results of our year-scale comparable with the Binet, we have continued the use of the 75 per cent criterion. To have used the 50 per cent criterion would have made the tests more difficult, because less age credit would have been assigned the successful completion of a design in the series.

The procedure in the use of this performance scale is the same as that for the regular Binet, except that for those unable to speak or understand verbal directions, gestures and imitation are the means utilized to present the problem. It is necessary, therefore, that the trial design should be utilized before entering upon the examination proper. For detailed instructions, however, the reader is referred to an earlier section (Part B, Chap. II, Section 2).

[1] "Some Logical Aspects of the Binet Scale," *Psychol. Rev.*, 1916, 23: 129–152, 165–179.

[2] "Further Logical Aspects of the Binet Scale," *Psychol. Rev.*, 1916, 23: 407–411.

2. THE PERCENTILE METHOD

Because of some important statistical fallacies underlying this method as now utilized — fallacies which only one or two have as yet pointed out [1] — we shall devote a rather extended discussion to this procedure.

The developments in the field of differential psychology have doomed forever that much popular slogan: "We are all born free and equal." The scientific study and analysis of individual differences has amply demonstrated that a given individual may rank first in one trait, last in another, and somewhere in between in another, and that this combination of ranks may be different for each individual. Thus in a group of one hundred, person A may rank 1st in intelligence, 10th in motor coördination, 85th in general reaction time, 100th in height, and so on. Person B may rank 12th in intelligence, 57th in motor coördination, 3d in general reaction time, 7th in height, and so on. Person C may present an entirely different array of traits. This principle has been utilized by the Russian psychologist Rossolimo, for the measurement and the graphic representation of developed mental processes, which he designates the "Psychological Profile Method." [2]

[1] See J. B. Miner: "Deficiency and Delinquency." Warwick and York, 1918; 355 pp. See esp. pp. 275–279.

S. C. Kohs: "Percentile Norms for Scaling Data," *J. of Educ. Psychol.*, 1918, 9: 101–102.

[2] G. Rossolimo: "Die psychologischen Profile. Zur Methodik der quantitativen Untersuchung der psychischen Vorgänge in normalen und pathologischen Fällen. Eine experimentell-psychologische Skizze." Klin. f. psych. u. nerv. Krankh., 1911, 6: 249–294. — English translation: B.

Traits may be compared either for their quantitative or qualitative differences. At the present time we have developed a finer methodology for the former than for the latter. Thus our comparisons of one individual with another, or of one individual with a group, are largely comparisons of a quantitative nature.

Theoretical Considerations:

The percentile method or the percentile system is a method for determining or locating the rank or the position of an individual within a group. All the members of a group are ranked with respect to a given trait, the lowest member at one end and the highest member at the other. This arrangement is illustrated in Graph 6 A, and the curve therefor has been called the "ogive" by Galton. Assume, for purposes of illustration, that we are studying the heights of 58 four-year-old boys. We arrange these measures in increasing order, the lowest is 866 mm., and the highest is 1165 mm. (See Smedley's Tables.) The curve connecting the separate heights of the 58 children is the ogive. Now if we run along the base from the 1st to the 6th case, the 6th case represents the 10th per cent case, and the height at this point is called the 10th percentile (964 mm.). Twice that distance along the base brings us to the 12th case, and the height at this point is called the 20th percentile (968 mm.). And three times the original distance brings us

Parker: "The Psychograph of Rossolimo," *Amer. J. of Insanity*, 1916, 73: 273–293.

See also Titchener Memorial Volume.

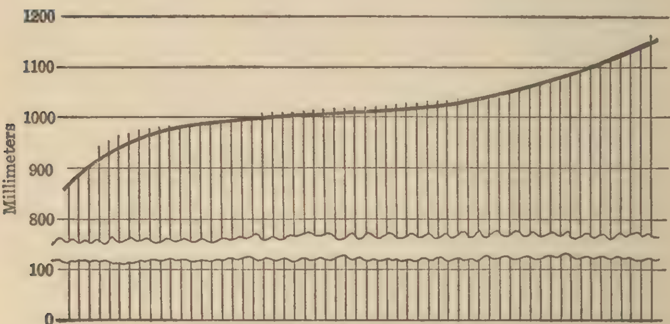
For criticism see Ed. Claparede: "Profiles psychologiques gradués d'après l'ordination des sujets." *Archives de Psychol.*, 1916, 16: 70–81.

to case 17, the height at this point, 997 mm., representing the 30th percentile. And so on, for the 40th, 50th, to the 100th percentile.

On the basis of this arrangement we may compare either two or more individuals for relative positions within the group, or we may compare the rank of one

GRAPH 6A

The Ogive



Heights of 58 four-year-old Boys arranged in order from lowest to highest. (After Smedley.)

in the light of the ranks of the other members of the group.

With this orientation let us turn to the question of tests. The last five years have seen a rapid multiplication of graded tests and scales, some intellectual, some pedagogical, some involving performance as distinct fom those tests involving the use of language symbols. With this growth has come the frequent suggestion that tests should be evaluated and interpreted in terms of

percentile norms, and that we utilize the percentile method for grading and comparing individuals. In fact, four of the most extensive studies made within recent years, by Smedley, by Woolley and Fischer, by Doll, and by Pintner and Paterson, have made use of just this principle.[1]

In the last instance particularly, has the value of the percentile method been emphasized:

> "The percentile method seems to offer the best possibilities for future work. The percentile divisions used can be made as small as the delicacy of the tests will warrant. This method is especially desirable because it permits us to compare an individual's performance with the performance of other individuals of the same age. It would seem at present, however, to require, for purpose of standardization, a very great number of unselected individuals at each age." (p. 212.)

A fundamental fallacy, however, underlies the present practice in the construction and the use of percentiles. The quite recent report that the percentile method is

[1] F. W. Smedley: "Report of the Department of Child-Study and Pedagogic Investigation. 1900–1901." Child Study Report No. 3. Chicago Public Schools, 1902.

H. T. Woolley and C. R. Fischer: "Mental and Physical Measurements of Working Children." *Psychol. Monogr.*, 1914, Vol. 18, No. 1. 247 p.

H. T. Woolley: "A New Scale of Mental and Physical Measurements for Adolescents and Some of its Uses." *J. of Educ. Psychol.*, 1915, 6: 521–550.

E. A. Doll: "Anthropometry as an Aid to Mental Diagnosis." Research Publication No. 8, Vineland, N. J. Training School, Feb. 1916. 91 p.

R. Pintner and D. G. Paterson: "A Scale of Performance Tests." New York, Appleton & Co., 1917. 218 p.

not meeting expectations cannot help but reinforce the writer's opinion that the source of the trouble lies perhaps in the weaknesses of our present technique. "The correlation of the percentile method with the Yerkes-Bridges is the lowest of the four methods and at present does not seem of much value. Possibly this is due to the comparatively small number of cases tested at each age."[1]

Among the important cautions to be observed in scale-making are two: First, be certain of the zero point; second, be sure that the units of the scale are equal. This latter is of vital importance where scores of various tests in a scale are totaled or averaged, and then evaluated in terms of percentile norms. All percentile differences should have a constant, not a variable connotation, and each percentile value should represent a definite and constant position on the normal frequency curve. This is far from true with our present technique.

Perhaps the chief and most destructive criticism of the present percentile method for scale construction is this: *Equal percentile differences* on the base of our orgive *do not represent equal differences in ability* between individuals. No one will argue that the amount of mental ability represented between percentiles 10 and 0, equals that between 20 and 10, equals that between 30 and 20, equals that between 40 and 30, equals that between 50 and 40, and so on. In fact, we know from the psychology of individual differences that the differences in ability

[1] Pintner, R., and Reamer, Jeanette C.: "Children Tested by the Point Scale and the Performance Scale." *Psychol. Clinic*, 1917, 11: 142–151. p. 150.

of adjacent cases among average, or mediocre, or most typical individuals is very small indeed, whereas the differences in ability of adjacent cases among extreme variates is, on the contrary, very great.

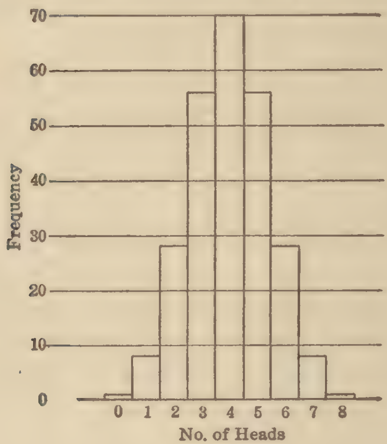
For a clearer elucidation of this problem we must turn to the actual normal probability curve.

The first clear explanation of the meaning and significance of the normal frequency curve came through the medium of an astronomer, Laplace — a mathematician, Gauss — and an anthropologist, Quetelet. The first two were engaged in efforts to determine the laws underlying the occurrence of purely accidental events, the last devoted a good part of his lifetime toward finding the "average man," lost somewhere midst a confusion of diversity. (It might not be amiss to mention here that just in that realm where one would least expect the operation of fixed laws, the realm of chance and accident, we actually do find a clear-cut subserviency to fundamental laws that definitely determine the frequency with which events occur.)

If we should toss eight pennies (perfectly symmetrical) some few million times, then once out of every 256 times they would all come down heads; 8 times out of every 256 times they would come down 7 heads and 1 tail; 28 times out of 256 they would come down 6 heads and 2 tails; 56 out of 256, 5 heads and 3 tails; 70 out of 256, 4 heads and 4 tails; and once out of every 256 times they would all come down tails (theoretically). If we plotted these data as a column diagram we would obtain Graph 7. Or plotted as a smooth frequency curve we could graphically represent the above facts as Graph 8.

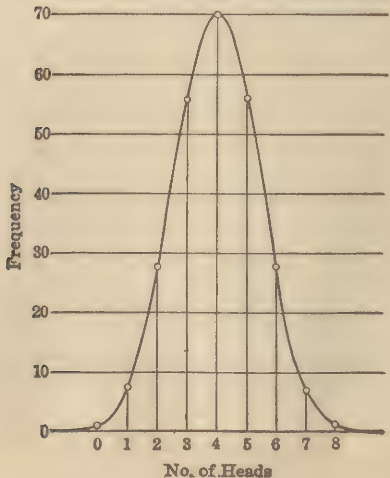
GRAPH 7

Theoretical Frequency
8 Pennies tossed 258 times



GRAPH 8

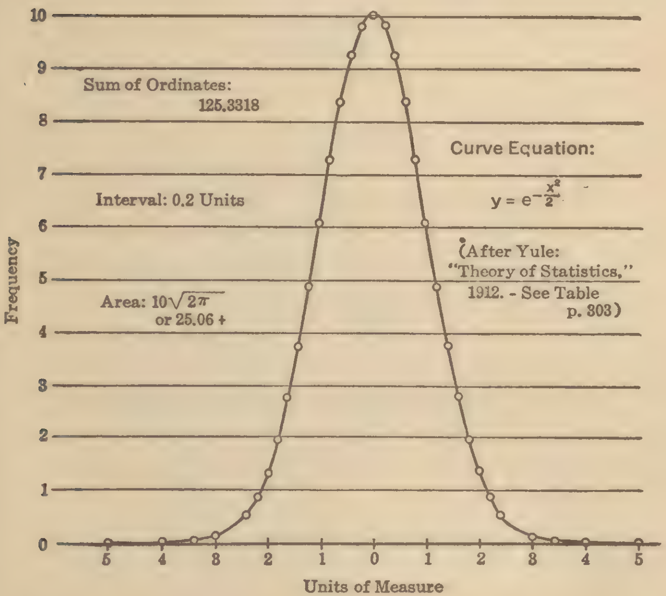
Theoretical Frequency
8 Pennies Tossed 258 Times



Laplace and also Gauss derived a formula for the theoretical normal probability curve which is given in Graph 9.

GRAPH 9

The Normal Curve



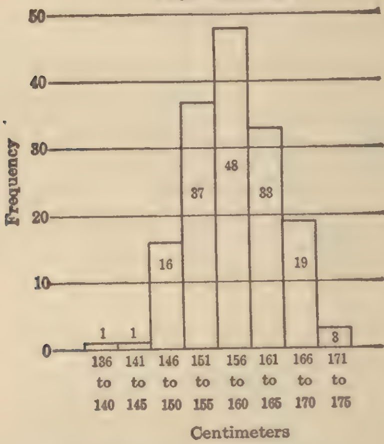
The probable occurrence of the different physical and mental traits possessed by human beings also distribute themselves, in many instances, in this bell-shaped fashion.[1] Thus, in Graphs 10, 11, and 12, are represented

[1] In this connection the writer wishes to call attention to an excellent article by E. G. Boring: "The Logic of the Normal Law of Error in Mental Measurement," *Amer. J. Psychol.*, 1920, 31: 1–33, dealing with

the frequencies of the heights of fifteen-year-old girls, fifth-grade ability in arithmetic and the differences in the intellectual ability of 905 unselected children examined with the Stanford-Binet Scale.

GRAPH 10

Heights of 158
15-year-old Girls



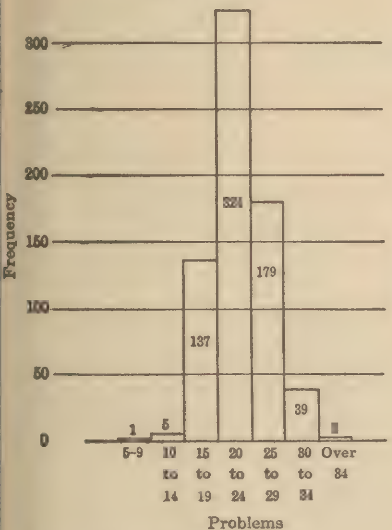
(from figures by B.T. Baldwin, p.15, "Physical Growth and School Progress", 1914).

For the purpose which we have set before ourselves at the beginning of this discussion, we cannot separate the normal probability curve from the ogive. Thus, for a given table of data, Table XXVII, Graph 13 is the ogive, and Graph 14 the frequency curve, both being plotted on the same coördinates.

this particular matter, in which he points out the fallacy of a too general application of the "normal curve" to measurements of mental phenomena.

GRAPH 11

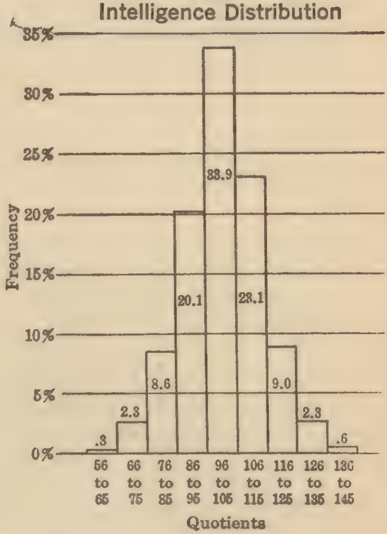
No. of Addition Problems
Solved by 687 Pupils
In Grade 5



(from figures by C. Woody, "Measurements of Some Achievements in Arithmetic," 1916, p.27).

GRAPH 12

905 Children
Intelligence Distribution



(from figures by Terman: "Measurement of Intelligence," 1916, p.68).

TABLE XXVII

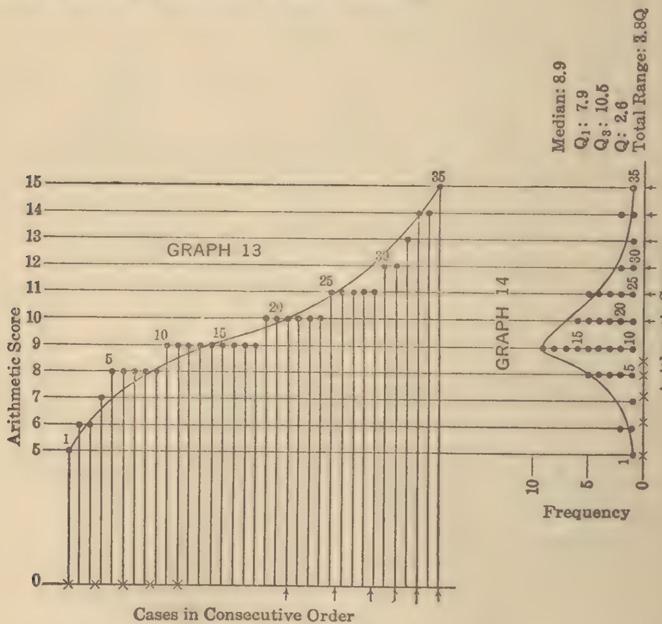
STARCH ARITHMETIC SCALE A *

Scores in a 7th Grade

SCORE	5	6	7	8	9	10	11	12	13	14	15	TOTAL
Frequency. . . .	1	2	1	5	9	6	5	2	1	2	1	35

* From Fig. 19, Starch: "Educational Psychology," 1920, p. 35.

In these diagrams items are numbered and their position on the ogive and on the column diagram may be clearly followed. It is of great importance to keep this relationship in mind, for in the confusion of the two curves lies the fallacy at the bottom of pretty nearly all the work in percentile scale making.



In Graphs 13 and 14 the crosses on the base of the ogive represent equal percentile (decile) differences between measures. These may be followed through to the base of frequency curve. On the other hand, the arrows represent equal units, decile or percentile, on the base of the frequency curve, these then are referred back to

ogive base. The marked lack of equality between the corresponding divisions is quite apparent. It should be emphasized that equal distances along the base of our ogive cannot, except under a combination of extremely unusual circumstances, yield equal distances along the base of our frequency curve. But scale values *must* be equal if we are going to equate, add, subtract, multiply, divide, or compare percentile scores. It is equally true that equal units along the base of our frequency curve will not yield corresponding equal units on the base of our ogive.

Another fact which requires mention is this: If we are going to determine the ability of a given individual, judged by reactions to more than one test, and utilizing the percentile method for determining his absolute score, then our percentile scale must be made up of score-values obtained from the base of our normal frequency curve, and *not* from the base of our ogive. The reason for this will be apparent from a study of the normal frequency distribution. It is evident that equal distances along the base of our normal curve represent *equal* differences in ability. This, unfortunately, is not true for the ogive. Consequently, if we take any percentile table in the current psychological or educational literature, ability represented by the difference between the 10th and 20th percentiles is far from equal to the ability represented by the difference between the 20th and 30th percentiles. But this *would* have been the case if the percentile scale had been laid off on the base of the normal frequency curve. These differences in ability represented by differences in consecutive percentiles *must* be equal, if we are

going to add, subtract, multiply, divide, equate, compare, percentile scores.

To sum up the remarks above: Percentile scales should be so constructed that the various steps or units represent equal differences in ability. For devising a percantile scale, the normal frequency curve should be utilized and not the ogive as the base for operations.

Practical Considerations

In the following section will be indicated the wide divergence of percentile-scale values obtained by present methods, and percentile-scale values obtained by suggested methods.

Example One

Smedley, in Child-Study Report No. 3 (1902), Chicago Public Schools, gives the following data for Grip of Left Hand, Fourteen-year-old Boys (p. 17):

TABLE XXVIII

Percentiles	100	90	80	70	60	50	40	30	20	10	0
L. Hand Grip	51	36	32	30	28	26	25	23	21	19	14

This table is merely chosen as typical. Any other might have been taken in his report to illustrate the conditions.

These percentiles were obtained in the following manner:

"The individual cards on which the measurements were recorded when the child was tested were arranged according to the size of the pupils in each measurement, grouped separately for each age in years. The minimum

measurement in each group gave the zero percentile for that group. To determine the ten percentile for that group, ten per cent of the number of cards was removed, beginning at the minimal end, and the highest measurement on the cards so removed was recorded as the desired ten percentile. Similarly the other percentiles were determined, the maximum measurement being recorded as the one hundred percentile." (P. 13.)

It is evident that the ogive rather than the normal frequency curve was used to determine percentile values. Since 359 cases were used by Smedley, it is a simple matter to reconstruct his initial table, which is here presented:

TABLE XXVIII A

LEFT HAND GRIP OF 359 BOYS

No. of Cases	L. Hand Grip	No. of Cases	L. Hand Grip	No. of Cases	L. Hand Grip	No. of Cases	L. Hand Grip
3	14	22	24	9	34	2	44
4	15	24	25	10	35	0	45
6	16	21	26	8	36	0	46
8	17	15	27	8	37	1	47
9	18	23	28	6	38	0	48
11	19	19	29	7	39	0	49
16	20	16	30	5	40	0	50
19	21	17	31	2	41	1	51
25	22	12	32	0	42		
20	23	10	33	0	43		

The point we have attempted to emphasize in this discussion is that in determining percentile differences among individuals, the curve of reference should be the frequency curve and not the ogive. The table given by Smedley is based on differences obtained from equal ogival distances. We herewith outline a procedure and present a table of percentile differences based on equal distances

along the base line of the frequency curve, given in terms of standard deviation units or quartile deviation (P. E.) units. To find a score, then, corresponding to a given percentile, 0, or 10, or 20, and so on, it is merely necessary to determine the case which falls nearest that percentile, then by interpolation to obtain the refined score of the case. That value, then, is the score corresponding to the desired percentile.

To illustrate: In the above example the range in score is from 14.17 to 51.50 (scores refined by interpolation), or a difference of 37.33 Kg. The percentiles then, in terms of this range, are as follows:

TABLE XXIX

RANGE DIVIDED INTO PERCENTILES

Percentile	Range Value	Percentile	Range Value
0	0.000	60	22.398
10	3.733	70	26.131
20	7.466	80	29.864
30	11.199	90	33.597
40	14.932	100	37.330
50	18.665		

Now, adding these separate values to the initial score, 14.17, the final step in the formation of the percentile table is taken:

TABLE XXX

SCORE VALUES OF PERCENTILES

Percentile	Score	Percentile	Score
0	14.170	60	36.568
10	17.903	70	40.301
20	21.636	80	44.034
30	25.369	90	47.767
40	29.102	100	51.500
50	32.835		

For purposes of comparison let us place in parallel position the two percentile tables, one obtained by the ogive method, the other by the frequency method, and then compare the scores.

TABLE XXXI

THE TWO SCORE VALUES CORRESPONDING TO THE SAME PERCENTILE COMPARED

Percentile	Score: Ogive Method	Score: Frequency Method	Percentile	Score: Ogive Method	Score: Frequency Method
0	14	14.170	60	28	36.568
10	19	17.903	70	30	40.301
20	21	21.636	80	32	44.034
30	23	25.369	90	36	47.767
40	25	29.102	100	51	51.500
50	26	32.835			

Let us also note the cases which locate the various percentiles by the two methods:

TABLE XXXII

THE CASES WHICH LOCATE THE PERCENTILE BY THE TWO METHODS

Percentile	Case Number: Ogive Method	Case Number: Frequency Method	Percentile	Case Number: Ogive Method	Case Number: Frequency Method
0	1	1	60	215	324
10	36	21	70	251	350
20	72	70	80	287	356
30	108	152	90	323	358
40	144	226	100	359	359
50	180	289			

The significance of this new percentile table[1] lies in the fact that the difference in ability between the first and 21st cases is equal to that between the 21st and 70th, the 152d and 226th, the 356th and 358th, or the 358th and 359th, and so on. By the ogive method percentiles are determined by grouping all of one's cases into ten *equally* numbered groups. By the frequency method, however, the number of cases between percentiles is markedly *unequal*, thus in our table the number of cases between consecutive percentiles is 20, 49, 82, 74, 43, 35, 26, 6, 2, 1. These numbers mean that it takes 2 individuals at one portion of the distribution to show as much difference in ability as is shown at other portions by 83, or at another by 20.

To make the differences between the two methods more apparent, the following two figures (Graphs 15

[1] In the above table for grips, the range in score is from 14.17 to 51.50, or 37.33. The arithmetic mean = 26.8 and the standard deviation is 6.46. Therefore, each unit of $\sigma = 5.78$ units of the range $\frac{37.33}{6.46}$. The total range, then, from the 0 percentile to any other is equal to the sigma corresponding to that percentile multiplied by 5.78. Since sigma = 6.46, the sigmas corresponding to the different percentiles are

TABLE XXXIII

SIGMAS CORRESPONDING TO VARIOUS PERCENTILES

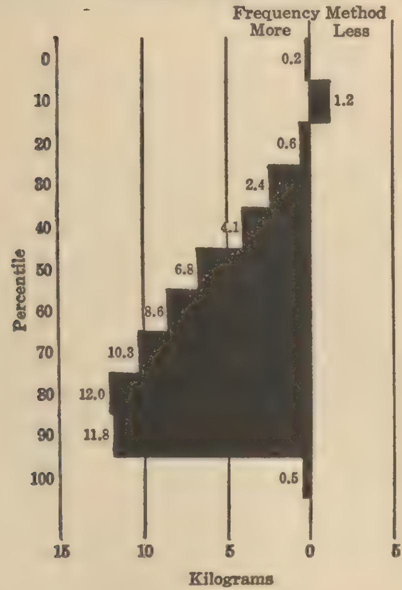
Percentile		Percentile	
0	0.000	60	3.876
10	0.646	70	4.522
20	1.292	80	5.168
30	1.938	90	5.814
40	2.584	100	6.460
50	3.230		

The same, of course, would hold true for Quartile (P. E.) values.

and 16) have been drawn to demonstrate first, the differences in score for each of the percentiles, and second, the differences in the positions of the individuals on the ogive who determine these percentile values.

GRAPH 15

Score Difference between Two Methods

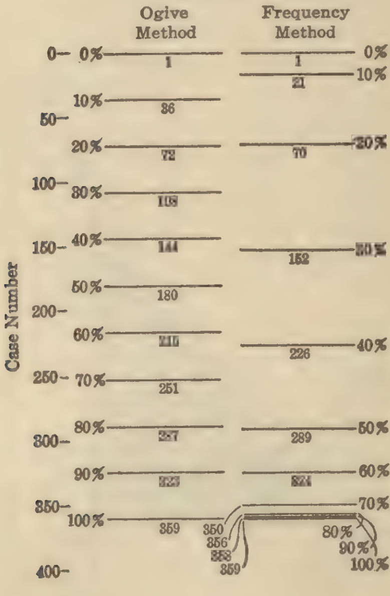


The block designs were not given percentile ratings, since it is an open question whether this method is an improvement either over the Year Scale Method or the P. E. Scale Method, both of which have been utilized in standardizing the block designs. In fact the percentile method correctly used reduces this standardiza-

tion method merely to a variation of the linear projection method which is discussed later

GRAPH 16

Case Locating Percentiles



3. THE POINT SCALE METHOD

A valuable discussion of this procedure (used by Yerkes, Bridges, and Hardwick) is found in Pintner and Paterson's "A Scale of Performance Tests" to which we have had occasion to refer in some earlier sections. They very clearly point out its shortcomings. The allotment of points is absolutely arbitrary. There is no guiding principle, the method is purely haphazard, and the

only check is the fear of making the mass of points cumbersome and unwieldy. Yet, that point scales are desirable is evidenced by the careful work of Otis (*Jour. of Educ. Psychol., 1918*). However, none of the point scales now in use are free from statistical and psychological objections which are inherent in the method and can only be removed by discarding entirely that method of standardization.

4. THE P. E. SCALE METHOD (Linear Projection Method)

This method has been used extensively at Columbia under the direction of Thorndike, and we have adopted that methodology with some modifications to a standardization of our block designs. The work of Buckingham in spelling, of Trabue with language scales, and of Woody with arithmetic scales, are examples. The chief advantage of this method of scaling tests is that each test in the series is given a definite weight depending on the successes or failures of children at each age. The normal probability curve is utilized in determining the location of a test on a P. E. Scale, the units of which are all equal. Each P. E. Scale has its zero point and each test is placed one above the other, depending upon the relative difficulty of each. By this means we have been able to project each design on a linear scale, each test being given a weight corresponding to its position on this scale. We were also interested in giving due weight to time and moves. For this purpose the times and moves for each design were correlated with life age. We found a most interesting situation, new to the standardization of mental tests. In correlating (*a*) suc-

cess or failure in the designs, (*b*) time, and (*c*) moves, with age, we discovered that with age (*a*) yielded a correlation coefficient of about 80, (*b*) about 40, and (*c*) about 20! In other words, the respective weights for success, time, and moves are 4:2:1. Therefore, when we obtained the P. E weights for each design, time and moves were each given their respective weights depending on the initial value of the design. (See Chap. IV.)

Our final standardization includes, therefore, a year scale such as the Binet, and a point scale derived by the P. E. scale method. The former may be supplemented by series of tests which others have standardized; the latter is a short scale complete in itself.

A word in conclusion. Periodically, statisticians engaged in standardizing mental or pedagogical tests should take stock of their methods, technique, and fundamental assumptions. It is an attitude of mind well worth cultivating not to regard as final any set procedure by means of which scales are at present derived. We are, unquestionably, inclined too frequently to forget that we are carrying over into education and psychology highly refined instruments which are of accepted service in biology, chemistry, and astronomy, but may not be utilized with the same degree of assurance and finesse in those sciences involving the human factor, such as sociology, psychology, and education. Well-merited criticism has already been launched against our blind utilization of correlation formulæ and of transmutation tables, and our incessant worship of the "normal curve" fetish.

In this chapter on standardization technique the writer is well aware of the many weaknesses in the methods for which preference is expressed. Until further progress is made, however, we must accept what we have with scientific caution, subjecting our procedures, at least, to the pragmatic test.

CHAPTER IV

The Standardization of the Block-Design Tests

1. *Introduction*
2. *Age Standardization*
   - (*a*) Mental Age Norms
   - (*b*) Sex Differences (Mental Age Norms)
   - (*c*) Chronological Age Norms
   - (*d*) Sex Differences (Chronological Age Norms)
   - (*e*) Graphic Presentation of Norms
   - (*f*) Final Year Scale
3. *P. E. Standardization*
   - (*a*) Method
   - (*b*) Time and Move Norms
   - (*c*) Result: The P. E. Scale
4. *Final Revision*
   - (*a*) Eliminations
   - (*b*) Sex Differences
   - (*c*) Final Norms

1. INTRODUCTION

THE block-design tests used for standardization consisted of a series of twenty designs, graduated in difficulty (see pp. 66–67) and a set of sixteen specially colored cubes which were given to the S with the request to put the blocks, four, or nine, or sixteen, together and have the design on the tops of the blocks when complete resemble

the design as indicated on the card set before him. In this manner we had hoped to carry over into performance tests Ebbinghaus's idea of a "*Kombinations-methode.*" In this we were doing no pioneer work, having been preceded by Healy, Knox, and Pintner, but this particular combination of method and apparatus is new. The cube material is described more in detail in Chapter II, Section 1.

Every one who has been given the test, from the youngest child to the adult, has manifested a keen interest in manipulating the material. There seems to be some fascination about performance tests in general which it would indeed be interesting to fathom. The directions which were given the S's have already been stated in full elsewhere. (See Chapter II, Section 2.) A record was kept of success or failure, enough time being allowed for success if that seemed at all possible, and our final results show, in fact, that we were too lenient to the extent of anywhere from a half-minute to a minute in the case of each design. The final time limits have all been reduced as a result of our findings. The original time limits for the designs, upon which success or failure mentioned in the succeeding tables was based, were:

Designs I–IX,	2½ minutes
Designs X–XI,	3 minutes
Designs XII–XVII,	4 minutes
Designs XVIII–XIX,	4½ minutes
Design XX	5 minutes

Moves and time were both recorded, the latter in whole seconds. These two factors, however, did not here affect the score: "passed" or "failed." For an explana-

tion of what was considered a move, see previous section on directions for applying the tests.

In the final standardization by the linear projection method three designs, XI, XIII, and XVII, were omitted and the others were assigned the following Arabic numerals:

Design	I	now Design	1	Design	IX	now Design	7
"	II	" "	2	"	X	" "	10
"	III	" "	3	"	XII	" "	11
"	IV	" "	4	"	XIV	" "	14
"	V	" "	5	"	XV	" "	12
"	VI	" "	6	"	XVI	" "	13
"	VII	" "	9	"	XVIII	" "	15
"	VIII	" "	8	"	XIX	" "	16

Design XX now Design 17

2. AGE STANDARDIZATION

We shall utilize both chronological and mental age standards in assigning age values to each of the designs. Some may object to the use of mental ages for standardization, but we have found no valid reasons for rejecting this criterion; in fact there are obvious reasons for its utilization. First, it serves as a check on an impartial selection of children — if they happen to be somewhat retarded or accelerated, the mental age difference will weight the norms accordingly. Secondly, we wish to translate our norms in terms of Binet mental age, so that a mental age obtained by this performance scale will possess the same significance as a Binet mental age. Nevertheless, Pintner and Paterson[1] maintain "Why should we know the mentality of the children we are

[1] "A Scale of Performance Tests," p. 52.

testing? What we want to arrive at is the ability of children of a specific age on a certain test, and we ought not to be influenced by other estimates of their ability. Our sole endeavor must be to get a fair sampling of cases at each age. If we standardize according to mental age arrived at by any scale, we are moving in a circle and presupposing that our determination of mentality is accurate and final." That we do not move in a circle, however, is clearly evident from the close correspondence between the chronological and mental age norms as indicated in the graphs. And the attempt to obtain a "fair" sampling at each age is rarely a statistical success.

(*a*) *Mental Age Norms*

We have utilized here the reactions of three different groups, (1) the whole P. S.[1] group, (2) the F. M.[2] and the P. S. groups, and (3) a group consisting only of children having I. Q.'s between 86 and 115. In this manner, the children of I. Q.'s between 86 and 115 have weighted our mental age norms to an extent of 3, the rest of the P. S. children, the accelerated and retarded, to an extent of 2, and the F. M. group to an extent of 1. This reduces any defect in the norms obtained because of the possible inaccuracy of the Binet scale in a few cases, this inaccuracy being more likely to occur with the so-called F. M., the seriously retarded, or the greatly accelerated.

Tables XXXIV, XXXV and XXXVI give the numbers and per cents passing for each of the sexes and as a total for the whole P. S. group. (Age 4 on the table includes ages between 3 years 7 months and 4 years 6 months, and so on for each age.)

[1] Public School. [2] Feeble-minded.

TABLE XXXIV

MENTAL AGE NORMS (P. S. GROUP)

Per Cent Passing (Boys)

Mental Age	No. of Cases	Design Number																			
		I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI	XVII	XVIII	XIX	XX
4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	6	33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	12	100	66	33	25	8	17	8	8	8	0	0	8	0	0	0	0	0	0	0	0
8	20	95	85	85	40	25	20	10	20	25	0	5	10	0	5	0	0	0	0	0	0
9	16	100	87	94	63	56	38	0	38	38	6	0	0	6	13	6	13	13	13	0	6
10	21	100	86	86	71	71	43	33	52	48	24	29	38	5	24	10	24	24	10	10	0
11	17	100	100	94	88	94	65	47	59	65	47	35	47	35	41	47	29	65	24	24	18
12	18	100	100	100	89	94	83	44	67	83	39	50	56	39	61	33	44	67	39	11	0
13	11	100	91	91	73	73	64	36	55	64	27	45	45	27	27	36	36	45	27	18	18
14	8	100	100	100	100	88	75	63	80	88	63	63	63	38	63	50	75	88	50	38	25
15	4	100	100	100	100	100	100	75	100	100	50	75	100	75	100	50	75	75	75	75	75
16	5	100	100	100	100	100	100	60	80	100	80	40	100	80	80	100	100	100	80	80	60
17	4	100	100	100	100	100	100	100	100	100	75	100	100	100	100	75	100	100	100	50	100
18	3	100	100	100	100	100	100	100	100	100	100	67	100	100	67	100	67	100	67	67	100
19	4	100	100	100	100	100	100	75	100	100	100	100	100	75	100	100	100	100	100	50	75

TABLE XXXV

Mental Age Norms (P. S. Group)

Per Cent Passing (Girls)

Mental Age	No. of Cases	Design Number I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI	XVII	XVIII	XIX	XX
6	13	69	38	38	15	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	12	75	25	33	8	17	8	8	8	0	0	0	0	0	0	0	0	0	0	0	0
8	22	95	73	64	41	27	22	9	22	32	5	5	0	0	0	0	0	0	0	0	0
9	13	100	69	85	31	38	38	8	15	15	8	8	8	8	0	0	0	0	0	0	0
10	17	100	100	76	94	76	53	24	59	47	6	24	41	12	6	12	24	18	18	0	0
11	20	100	100	95	85	75	75	30	60	65	25	30	30	15	10	15	20	30	20	5	0
12	10	100	80	90	90	80	70	50	60	60	20	10	30	10	30	50	40	60	20	0	0
13	7	100	100	100	71	71	86	43	71	57	29	57	57	29	43	43	57	57	29	0	14
14	8	100	100	100	100	100	88	63	88	100	25	63	63	25	63	50	38	75	50	38	25
15	6	100	100	100	100	100	100	100	83	83	67	50	83	17	50	50	67	67	67	33	33
16	5	100	100	100	100	100	100	60	80	100	80	60	60	80	80	80	80	80	80	60	40
17	3	100	100	100	100	100	100	100	100	100	100	67	100	100	100	100	100	100	100	67	67
18	1	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19	1	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

TABLE XXXVI

Mental Age Norms (P. S. Group)

Per Cent Passing (Total)

Mental Age	No. of Cases	Design Number																			
		I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI	XVII	XVIII	XIX	XX
4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	[illegible]	0	0
6	19	58	26	26	10	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	24	88	46	33	17	13	13	8	8	4	0	0	4	0	0	0	0	0	0	0	0
8	42	[illegible]5	79	74	41	26	21	10	21	29	2	5	5	0	2	0	0	0	0	0	0
9	29	100	79	90	48	48	38	4	28	28	7	4	4	7	7	4	7	7	7	0	4
10	38	100	92	82	82	74	47	29	56	47	16	26	39	8	16	11	24	21	13	5	0
11	37	100	100	95	86	84	70	38	59	65	35	32	38	24	24	30	24	46	22	14	8
12	28	100	93	96	89	89	79	46	64	75	32	36	46	29	50	39	43	64	32	7	0
13	18	100	94	94	72	72	72	39	61	61	28	50	50	28	33	39	44	50	28	11	17
14	16	100	100	100	100	94	81	63	88	94	44	63	63	31	63	50	56	81	50	38	25
15	10	100	100	100	100	100	100	90	90	90	60	60	90	40	70	50	70	70	70	50	50
16	10	100	100	100	100	100	100	60	80	100	80	50	80	80	80	90	90	90	80	70	50
17	7	100	100	100	100	100	100	100	100	100	86	86	100	100	100	86	100	100	100	57	86
18	4	100	100	100	100	100	100	100	100	100	100	75	100	100	75	100	75	100	75	75	100
19	5	100	100	100	100	100	100	80	100	100	100	100	100	80	100	100	100	100	100	60	80

TABLE XXXVII

MENTAL AGE NORMS (P. S. GROUP) (I. Q. BETWEEN 86 AND 115)

Per Cent Passing

Mental Age	No. of Cases	Design Number																			
		I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI	XVII	XVIII	XIX	XX
5	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	16	50	18	13	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	16	88	56	38	25	6	13	6	6	6	0	0	6	0	0	0	0	0	0	0	0
8	27	96	44	70	41	26	22	7	19	30	3	7	3	0	3	0	0	0	0	0	0
9	20	100	80	90	35	35	25	0	5	10	10	0	0	0	5	5	5	5	5	0	0
10	23	100	91	83	83	87	61	22	61	52	22	22	52	9	22	13	26	30	13	9	0
11	24	100	100	92	88	92	71	46	58	67	33	38	42	21	29	29	21	46	21	8	8
12	16	100	100	100	88	94	88	63	75	94	25	44	56	31	63	50	56	81	44	13	0
13	12	100	92	92	75	75	83	42	67	67	25	50	50	25	33	42	42	50	25	25	17
14	12	100	100	100	100	92	75	58	83	92	42	58	50	33	50	42	50	75	58	25	25
15	9	100	100	100	100	100	100	89	89	89	56	56	89	44	78	44	67	67	67	56	56
16	7	100	100	100	100	100	100	43	71	100	71	29	71	71	71	86	86	86	86	57	29
17	4	100	100	100	100	100	100	100	100	100	75	75	100	100	100	100	100	100	100	75	75
18	1	100	100	100	100	100	100	100	100	100	100	0	100	100	0	100	100	100	100	100	100

TABLE XXXVIII

MENTAL AGE NORMS (F. M. AND TOTAL P. S. COMBINED)

Per Cent Passing

Mental Age	No. of Cases	Design Number																			
		I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI	XVII	XVIII	XIX	XX
3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	31	55	23	19	0	0	32	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	29	86	48	38	21	14	18	7	7	4	0	0	4	0	0	0	0	0	0	0	0
8	51	90	76	71	37	27	20	10	18	25	2	4	4	0	2	0	0	0	0	0	0
9	37	97	78	86	49	43	38	3	24	24	8	5	5	5	8	5	8	8	5	0	3
10	44	98	86	82	73	68	45	25	50	41	14	23	34	7	14	9	20	18	11	5	0
11	44	100	95	91	80	80	65	39	55	59	32	30	36	23	25	30	23	43	23	14	9
12	31	100	94	97	87	90	81	45	65	74	32	35	45	26	52	42	45	61	29	7	0
13	20	100	95	95	75	75	70	35	60	60	25	45	45	25	30	35	40	45	25	10	15
14	17	100	100	94	100	88	76	59	82	88	41	59	50	20	59	47	53	76	47	35	24
15	11	100	100	100	100	100	91	86	91	86	55	55	86	36	64	45	64	64	64	45	45
16	11	100	100	100	100	100	100	64	73	100	73	55	86	86	73	91	91	91	86	64	55
17	7	100	100	100	100	100	100	100	100	100	86	86	100	100	100	86	100	100	100	57	86
18	4	100	100	100	100	100	100	100	100	100	100	75	100	100	75	100	75	100	75	75	100
19	5	100	100	100	100	100	100	80	100	100	100	100	100	80	100	100	100	100	100	60	80

TABLE XXXIX

MENTAL AGE NORMS (COMPOSITE)

Per Cent Passing

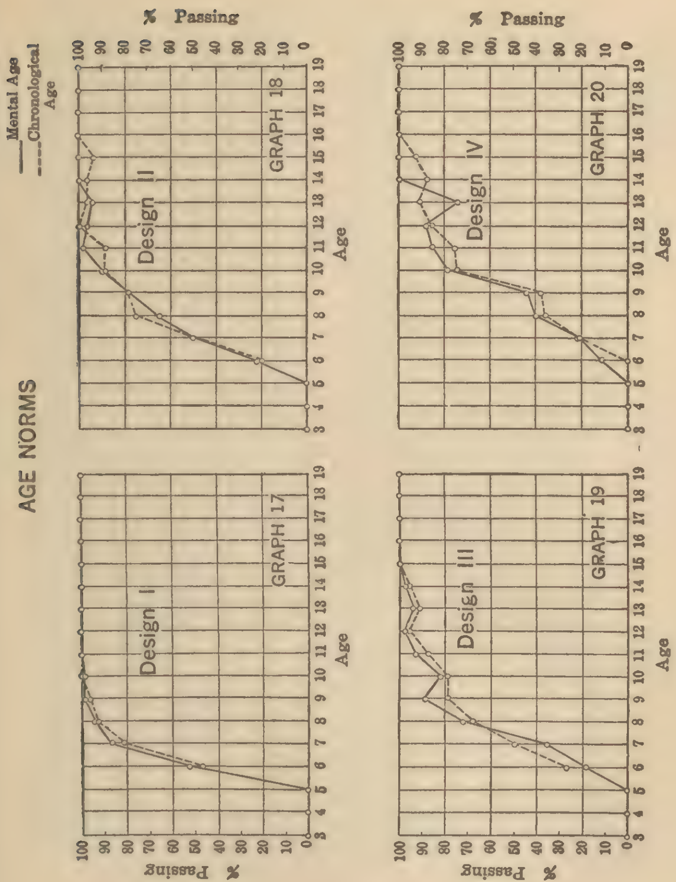
Mental Age	Design Number																			
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI	XVII	XVIII	XIX	XX
3 . . .	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4 . . .	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5 . . .	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6 . . .	53	22	19	12	0	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7 . . .	87	50	36	21	11	15	7	7	5	0	0	5	0	0	0	0	0	0	0	0
8 . . .	94	66	72	40	26	21	9	19	28	2	5	4	0	2	0	0	0	0	0	0
9 . . .	99	79	89	44	42	34	2	19	21	8	3	3	4	7	5	7	7	6	0	2
10 . . .	99	90	82	79	76	51	25	56	47	17	23	42	8	17	11	23	23	12	6	0
11 . . .	100	98	93	85	85	69	44	57	64	33	33	39	23	26	30	23	45	22	12	8
12 . . .	100	96	98	88	91	83	51	68	81	30	38	49	29	55	44	48	69	35	9	0
13 . . .	100	94	94	74	74	75	39	63	63	26	48	48	26	32	39	42	48	26	15	17
14 . . .	100	100	98	100	91	77	60	84	91	43	60	57	31	57	46	53	77	52	33	25
15 . . .	100	100	100	100	100	100	88	90	88	57	57	88	40	71	46	67	67	67	50	50
16 . . .	100	100	100	100	100	100	56	75	100	75	45	79	79	75	89	89	89	84	67	45
17 . . .	100	100	100	100	100	100	100	100	100	82	82	100	100	100	91	100	100	100	63	82
18 . . .	100	100	100	100	100	100	100	100	100	100	50	100	100	50	100	83	100	83	83	100
19 . . .	100	100	100	100	100	100	80	100	100	100	100	100	80	100	100	100	100	100	60	80

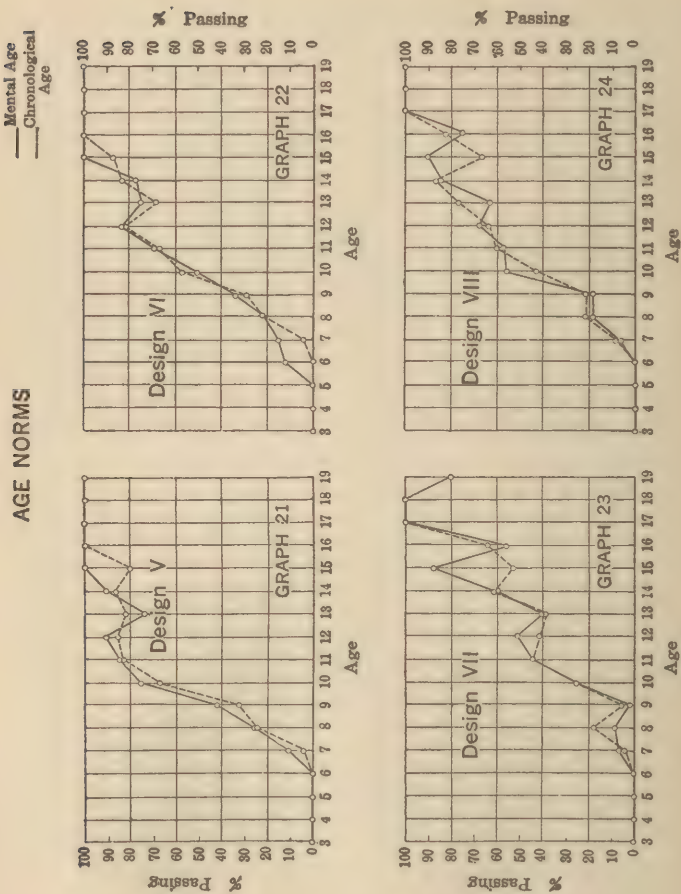
Table XXXVII presents the data for the per cents passing at each mental age, the group here being limited to those only in the P. S. group whose I. Q.'s ranged between 86 and 115.

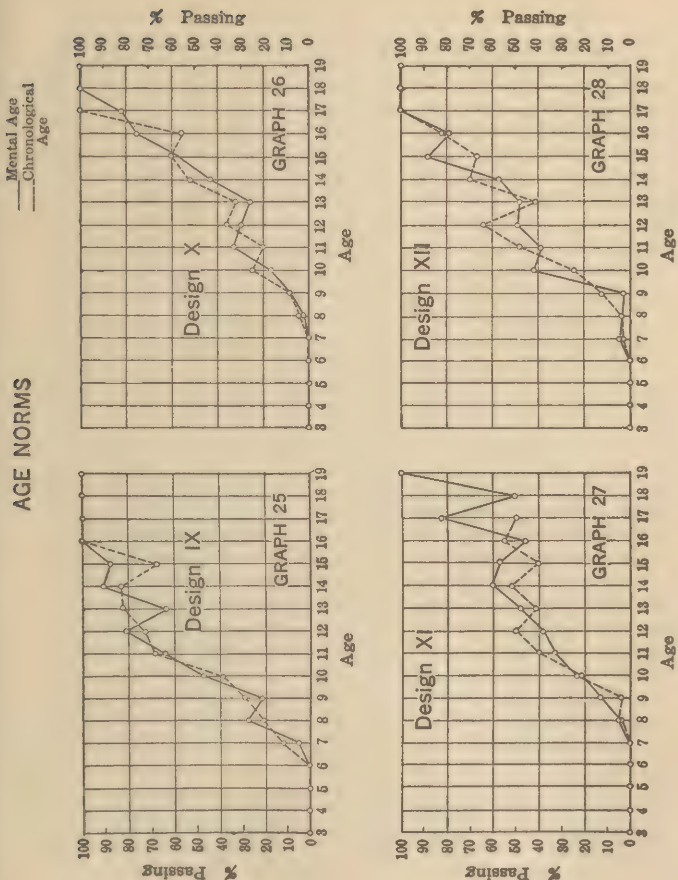
Table XXXVIII gives the per cents passing at each mental age, combining the F. M. with the total P. S. group.

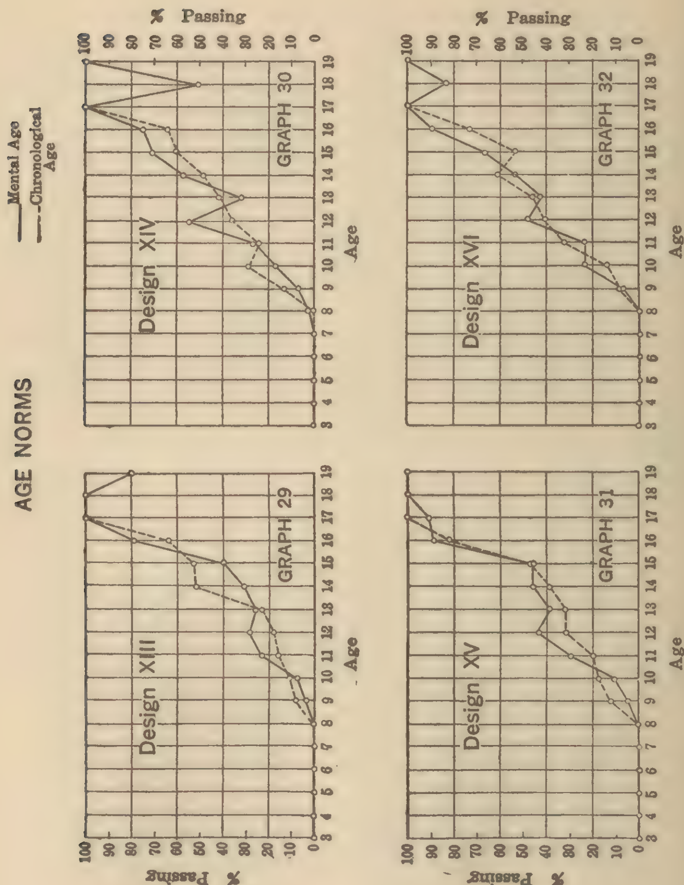
In Table XXXIX is presented a composite of the data in Tables XXXVI, XXXVII and XXXVIII, the arithmetic mean being used. It has already been mentioned that as a result of this procedure the I. Q. norms receive a weight of 3, the norms for the P. S. children under 86 I. Q. and above 115 I. Q. a weight of 2, and the norms of the F. M. group a weight of 1. The figures for each of the designs in this table were utilized for the final mental age graphs showing per cents passing at each mental age (Graphs 17 to 36).

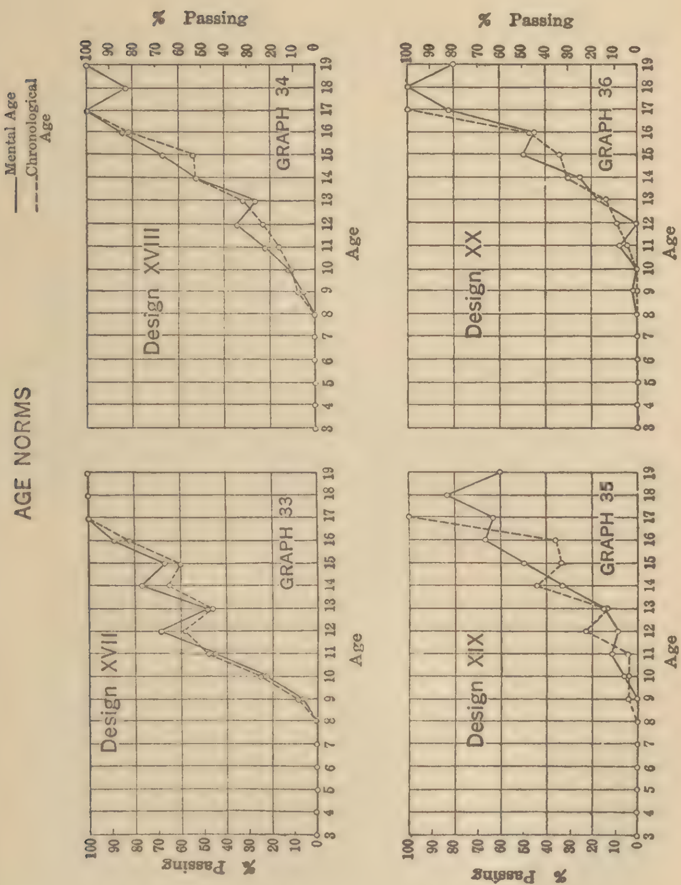
A matter which should be observed in connection with the composite norms is the fact that the children falling in the 13-year (mental age) group, are the only ones which do not manifest a consistent superiority over the children of the next preceding age. It is true that in many instances the difference is no more than 3 or 4 per cent in either direction. But not only should this slight negative difference not exist, but the 13-year-olds should show marked superiority over the 12-year-olds. Biased as all devisers of tests are in favor of their own scheme of mental examination, the writer might now justifiably follow the custom of throwing the blame on the Binet tests. In this attempt, however, he is handicapped by too many years' actual experience with the Scale, during


AGE NORMS
Mental Age
Chronological Age
% Passing
Age
Design I
GRAPH 17
Design II
GRAPH 18
Design III
GRAPH 19
Design IV
GRAPH 20


AGE NORMS
Mental Age
Chronological Age
Design V
GRAPH 21
Design VI
GRAPH 22
Design VII
GRAPH 23
Design VIII
GRAPH 24
% Passing
Age


AGE NORMS
Mental Age
Chronological Age
Design IX
GRAPH 25
Design X
GRAPH 26
Design XI
GRAPH 27
Design XII
GRAPH 28
% Passing
Age


AGE NORMS
Mental Age
Chronological Age
Design XIII
GRAPH 29
Design XIV
GRAPH 30
Design XV
GRAPH 31
Design XVI
GRAPH 32
% Passing
Age


AGE NORMS
Mental Age
Chronological Age
Design XVII
GRAPH 33
Design XVIII
GRAPH 34
Design XIX
GRAPH 35
Design XX
GRAPH 36
% Passing
Age

which period he has developed no little regard for its accuracy. The same inadequacy of the 13-year-old group, though to a less marked degree, is observable in the table giving the chronological age norms.

(b) Sex Differences (Mental Age Norms)

Table XL summarizes the data for sex differences at ages 6–8, 9–13, and 14–19. This division is merely one of convenience, dividing the range of mental ages into approximately three equal parts. In this table are presented the number of times at the various ages for all of the designs that the boys show a greater percentage of passes than the girls, in the next row the number of times that this is true for girls, and the last row the number of times when the per cent passing is equal for the sexes. This number of equal per cents passing is great because 0 per cents and 100 per cents passing have been included.

TABLE XL

SEX DIFFERENCES

	Mental Ages			Total
	6–8	9–13	14–19	
Number of instances in which per cent of boys passing is greater than girls	11	53	23	87
Number of instances in which per cent of girls passing is greater than boys	13	35	18	66
Instances of no difference in per cent passing between boys and girls	36	12	79	127
Total	60	100	120	280

The data of Table XL is summarized more briefly and for per cents in Table XLI. It will be evident that no marked sex differences are observable here, the boys appearing somewhat more superior than the girls. This will be corroborated again at a number of other points.

TABLE XLI

SEX DIFFERENCES

	Number	Percentage of Total
Instances boys superior to girls . . .	87	31.1
Instances girls superior to boys . . .	66	23.6
Equality	127	45.3
Total	280	100.0

Table XLII presents this evidence again for the first ten designs, and for the last ten, and for the total.

TABLE XLII

SEX DIFFERENCES

Design No.	Mental Ages						Total	
	6–8		9–13		14–19			
	I–X	XI–XX	I–X	XI–XX	I–X	XI–XX	I–X	XI–XX
Number of instances boys superior . .	8	3	24	29	3	20	35	52
Per cent . .	27%	10%	48%	58%	5%	33%	25%	37%
Number of instances girls superior . .	13	0	18	17	7	11	38	28
Per cent . .	43%	0%	36%	34%	12%	18%	27%	20%
Number of instances equal	9	27	8	4	50	29	67	60
Per cent . .	30%	90%	16%	8%	83%	49%	48%	43%
Total	30	30	50	50	60	60	140	140
Per cent . .	100%	100%	100%	100%	100%	100%	100%	100%

In designs I–X, ages 6–8, girls seem superior, with no significant sex difference on the last ten designs. Between the ages 9–13, boys seem superior throughout. Between the ages 14–19 there is a good deal of equality, though girls are somewhat superior on the first ten and the boys on the last ten designs. In designs XI–XX, boys are superior at each of the grouped ages. On designs I–X girls are superior at each of the grouped ages except 9–13. From the table it is clear that in the early ages, from 6–8, there is no observable sex difference, perhaps girls may be said to be slightly superior, from 9–13 boys are somewhat superior, and from 14 to 19 boys are only slightly superior, with marked equality between the sexes at this age group.

To summarize the data for the sexes: In designs I–X, ages 6–8, girls are superior; ages 9–13, boys are superior; ages 14–18, girls are again superior. In designs XI–XX boys are superior at every age group. In the first ten designs the girls are superior, in the last ten the boys are superior. Perhaps the mechanical, somewhat puzzle-like nature of the later designs may help account for this difference. Taking the data in Table XL for *all* the designs, there is no significant sex difference from ages 5–8; between 9–13 and 14–18 boys seem superior. Combining all the tests and all the ages, the boys are slightly superior. It should be understood by the reader, however, that these tables merely show tendencies and are not to be interpreted as conclusive evidence for any valid argument relative to intelligence differences between the sexes.

(*c*) *Chronological Age Norms*

In obtaining these values we utilized only the P. S. cases, discarding all those who were not in the normal grade for their chronological age, or not within one year of normal grade. This yielded some 241 cases (121 boys, 120 girls) upon which these norms are based.

In Tables XLIII, XLIV, and XLV are presented the number and per cents of boys, of girls, and of the total number at each age passing each of the designs. The range in chronological age is smaller than the range in mental age, as would naturally be expected, the lowest chronological age falling in the 6-year group and the highest in the 17-year group. This yields a range of twelve years.

Attention should be called to the striking similarity between the tables for mental age and the succeeding tables for chronological age. There are no marked discrepancies in spite of the fact that the mental age tables included reactions from the feeble-minded, which would theoretically produce some disturbing effects.

The adequacy of the sampling and the validity of the standardization is also positively indicated by the close agreement between the respective tables for mental age and chronological age. Although the standardization technique for chronological age compels a reduction in the number of cases under consideration to 241, an elimination amounting to thirty-three per cent, nevertheless the figures obtained from this smaller and, in some respects, more homogeneous group, agree very closely with the figures of the preceding section, based upon mental age and on the total 366 cases.

TABLE XLIII

CHRONOLOGICAL AGE NORMS (AT OR WITHIN 1 YR. OF GRADE)

Per Cent Passing (Boys)

Life Age	No. of Cases	Design Number																			
		I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI	XVII	XVIII	XIX	XX
6	6	33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	14	86	64	57	21	7	0	7	7	14	0	0	7	0	0	0	0	0	0	0	0
8	10	80	80	60	30	10	20	20	20	10	0	0	10	0	0	0	0	0	0	0	0
9	13	100	85	85	46	39	31	0	23	31	8	8	23	8	15	15	8	8	8	8	0
10	16	100	88	88	75	63	50	38	44	50	31	38	31	13	44	25	19	38	13	6	0
11	15	100	87	87	73	87	67	40	67	67	27	40	40	27	27	13	33	47	20	7	7
12	10	100	100	90	80	90	70	50	70	80	60	40	70	0	60	40	40	60	10	30	10
13	9	100	100	100	89	89	67	44	89	100	33	56	44	33	44	44	56	44	22	11	0
14	14	100	93	93	86	86	79	57	86	86	50	71	79	57	57	36	71	79	57	50	36
15	7	100	100	100	86	86	86	43	57	57	57	29	71	57	71	57	57	71	57	43	43
16	5	100	100	100	100	100	100	80	80	100	60	60	100	80	60	100	100	100	100	40	80
17	2	100	100	100	100	100	100	100	100	100	100	50	100	100	100	100	100	100	100	100	100

TABLE XLIV

CHRONOLOGICAL AGE NORMS (AT OR WITHIN 1 YR. OF GRADE)

Per Cent Passing (Girls)

Life Age	No. of Cases	Design Number																			
		I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI	XVII	XVIII	XIX	XX
6	9	56	33	44	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	12	75	33	42	25	0	8	0	8	8	0	0	0	0	0	0	0	0	0	0	0
8	18	100	72	72	49	33	22	17	22	28	6	6	0	0	0	0	0	0	0	0	0
9	11	91	73	73	27	27	27	9	18	27	9	18	0	9	9	9	9	9	9	0	0
10	12	100	92	67	75	75	67	8	42	25	17	0	17	8	8	8	8	8	8	0	0
11	10	100	90	90	80	80	70	50	50	70	10	40	60	0	10	30	30	50	10	0	0
12	12	100	100	100	92	83	92	33	58	67	17	58	58	33	17	25	42	58	33	17	8
13	13	100	92	85	92	77	69	39	69	60	31	31	39	15	39	23	39	46	39	15	23
14	9	100	100	100	89	89	89	67	89	78	56	22	56	44	33	44	44	44	44	33	22
15	8	100	88	100	100	75	88	63	75	75	63	50	63	50	50	38	50	50	50	25	25
16	6	100	100	100	100	100	100	50	63	100	50	50	67	50	67	67	50	67	67	33	17

TABLE XLV

Chronological Age Norms (at or within 1 Yr. of Grade)

Per Cent Passing (Total)

Life Age	No. of Cases	Design Number																			
		I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI	XVII	XVIII	XIX	XX
6	15	47	20	27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	26	81	50	50	23	4	4	4	8	12	0	0	4	0	0	0	0	0	0	0	0
8	28	93	75	68	36	25	21	18	21	21	4	4	4	0	0	0	0	0	0	0	0
9	24	96	79	79	38	33	29	4	21	29	8	13	13	8	13	13	8	8	8	4	0
10	28	100	89	79	75	68	57	25	43	39	25	21	25	11	29	18	14	25	11	4	0
11	25	100	88	88	76	84	68	44	60	68	20	40	48	16	24	20	32	48	16	4	4
12	22	100	100	96	86	86	82	41	64	73	36	50	64	18	36	32	41	59	23	23	9
13	22	100	96	91	91	82	68	41	77	82	32	41	41	23	41	32	46	46	32	14	14
14	23	100	96	96	87	87	83	61	87	83	52	52	70	52	48	39	61	65	52	44	30
15	15	100	93	100	93	80	87	53	67	67	60	40	67	53	60	47	53	60	53	33	33
16	11	100	100	100	100	100	100	64	82	100	55	55	82	64	64	82	73	82	82	36	46
17	2	100	100	100	100	100	100	100	100	100	100	50	100	100	100	100	100	100	100	100	100

(d) Sex Differences (Chronological Age Norms)

Table XLVI analyzes the sex differences in the different age groups for the first ten designs, for the last ten, and for the total.

TABLE XLVI

SEX DIFFERENCES

Design No.	Chronological Age						Total	
	6–8		9–13		14–17			
	I–X	XI–XX	I–X	XI–XX	I–X	XI–XX	I–X	XI–XX
Number of instances boys superior	8	2	28	31	6	27	42	60
Per cent	27%	7%	56%	62%	20%	90%	38%	55%
Number of instances girls superior	14	1	16	16	14	3	44	20
Per cent	46%	3%	32%	32%	47%	10%	40%	18%
Number of instances equal	8	27	6	3	10	0	24	30
Per cent	27%	90%	12%	6%	33%	0%	22%	27%
Total	30	30	50	50	30	30	110	110
Per cent	100%	100%	100%	100%	100%	100%	100%	100%

It will be observed that girls are superior in designs I–X at ages 6–8 and 14–16, the boys being superior at ages 9–13. On the whole, for the first ten designs the girls are slightly superior. The boys are superior on designs XI–XX at all the age groups. Taking all the designs together, girls are superior at the earlier ages, boys becoming superior at about nine and remaining so thereafter. On the whole, it appears that boys are superior in these designs to the girls. These results based on chronological age confirm the findings by mental age.

Further analyses for sex differences have been made by means of a tabular comparison of results obtained (*a*) for the mental age (total P. S.) group, (*b*) for the mental age (selected I. Q.'s) group, and (*c*) for the chronological age group, by years below 9, between 9 and 13, and above 13. These data are not here presented, but may be constructed from the tables already given. It is evident (*a*) that for ages below 9, girls are slightly superior to the boys, this being especially true for the first ten designs. For the next ten designs there is no appreciable difference, the same being true for all the designs for children below nine years of age. (*b*) For ages 9–13, boys are superior to girls on all the designs and much more so on the second set of ten than on the first set. And (*c*) above age 13, boys seem slightly superior to girls, but on the first ten designs girls are superior to boys, while on the last ten, boys are quite superior to girls.

Considering merely designs I–X, and allowing for errors of sampling, we may say that there are no significant differences between boys and girls, but as regards designs XI–XX, with increase in age, boys seem superior to girls. Whether this difference is due to a difference in analytic synthetic power, or whether it may be largely accounted for by specialized mechanics on the part of boys, cannot be stated dogmatically. Then, again, whether this evidence would change with the addition of further samplings cannot be definitely foretold.

(*e*) *Graphic Presentation of Norms*

With the data then of Tables XXXIX and XLV we may plot our curves for per cent passing with increase

of mental and of chronological age. This has been done in Graphs 17 to 36. *It should be noted that the mental age and chronological age curves overlap, and if no key had been presented stating which is which, no one would have been able to tell them apart.* Many have been the warnings not to standardize a test on the basis of mental age, nevertheless the results are practically identical.

(*f*) *Final Year Scale*

Our final year-location of a test has been determined by combining (averaging) the two curves, mental age and chronological age. See Table L. Using the 50 and 75 per cent criteria, the designs are located at the ages stated in Table XLVII. Table XLVIII presents this material in another form. And Table XLIX gives the scores in designs successfully completed at the various age levels. Graphs 37 and 38 present the data graphically, and we have here added in Graph 39, for purposes

TABLE XLVII

YEAR SCALE NORMS

(Based on a composite of mental and chronological age norms)

Design	Age Value		Design	Age Value	
	50% Criterion	75% Criterion		50% Criterion	75% Criterion
	Yrs. Mo	Yrs. Mo		Yrs. Mo	Yrs. Mo
I	6–0	6–9	XI	13–6	18–0
II	7–0	8–6	XII	13–0	15–3
III	7–3	8–3	XIII	15–0	16–3
IV	9–3	10–6	XIV	13–6	16–0
V	9–3	10–3	XV	15–0	16–0
VI	10–0	12–6	XVI	13–3	15–9
VII	13–0	16–0	XVII	11–3	15–0
VIII	10–3	13–3	XVIII	14–0	15–6
IX	10–3	12–9	XIX	15–6	17–6
X	14–0	16–9	XX	16–0	17–3

TABLE XLVIII

YEAR SCALE NORMS

AGE	DESIGN NUMBER	
	50% CRITERION	75% CRITERION
6	I	I
7	II III	
8		II III
9	IV V	
10	VI VIII IX	IV V
11	XVII	
12		VI IX
13	VII XI XII XIV XVI	VIII
14	X XVIII	
15	XIII XV XIX	XII XVI XVII XVIII
16	XX	VII X XIII XIV XV
17		XIX XX
18		XI

TABLE XLIX

SCORES TO BE EXPECTED AT THE VARIOUS AGE LEVELS

AGE	(50%) SCORE	(75%) SCORE	AGE	(50%) SCORE	(75%) SCORE
4	0	0	12	9	7
5	0	0	13	14	8
6	1	1	14	16	8
7	3	1	15	19	12
8	3	3	16	20	17
9	5	3	17	20	19
10	8	5	18	20	20
11	9	5			

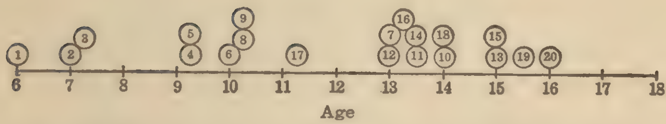
of comparison, the location of these designs on a P. E. Scale, the derivation of which is explained in the division on "P. E. Scale Norms."[1]

[1] It should be noted that although the original time limits for each design were somewhat longer than the later revision, the revision itself covers practically the same amount of time in which success may reasonably be expected. The validity of the above-indicated norms are, therefore, in no way affected by this change.

GRAPH 37

Age Location of Designs

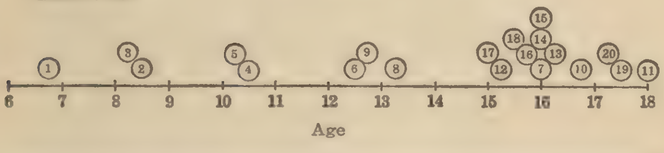
(50% Criterion)



Age Location of Designs

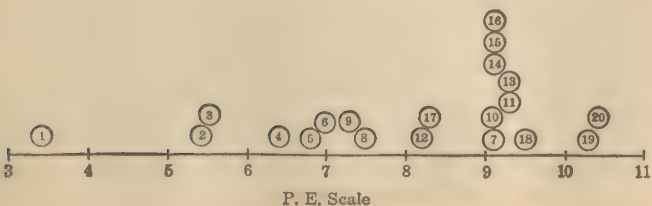
(75% Criterion)

GRAPH 38



GRAPH 39

P. E. Location of Designs



3. P. E. STANDARDIZATION

(*a*) *Method*

For a full and detailed explanation of the principles underlying the scaling of tests on a P. E. basis and for the technique involved, the reader is referred to the work of Buckingham, Trabue, and Woody. Here, let it suffice to mention that by means of this method, each test in the series is weighted, first, depending on the P. E. intervals between the separate ages, and second, depending on the ratios between the number of passes and failures. In Table L are presented the per cents passing each design at the various ages from 3 to 19. (This data is based on the age norms, mental and chronological, smoothed.) In Table LI are presented the differences between 50 per cent and the per cent passing a design at each age. The data are thus prepared for the use of Sheppard's Tables,[1] and in Table LII are accordingly presented the P. E. equivalents of the items in Table LI. In Table LIII are presented the P. E. intervals between consecutive ages, the bottom of the table showing the average intervals between consecutive ages. In Table LIV are tabulated the distances of each age median above the zero of the P. E. scale. The separate age-distributions are presented graphically in Graph 40.

It should be stated that a few minor modifications have been made in the application of this method to our problem. These changes have been in the direction of eliminating some of the super-refinements. In theory and result the method has been kept intact

[1] See Appendix III.

TABLE L

Per Cent Passing Designs

(Based on Age Norms, Mental and Chronological, — Smoothed)

Designs (No.)	Age 3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
I	0	0	0	50	83	93	98	100	100	100	100	100	100	100	100	100	100
II	0	0	0	21	49	65	78	89	94	96	97	98	99	100	100	100	100
III	0	0	0	19	46	68	80	86	91	94	96	97	99	100	100	100	100
IV	0	0	0	6	20	34	50	68	79	85	88	93	97	100	100	100	100
V	0	0	0	0	10	25	40	70	79	83	86	88	94	100	100	100	100
VI	0	0	0	4	11	20	33	50	66	73	78	84	93	100	100	100	100
VII	0	0	0	0	4	8	13	20	29	38	48	57	65	75	84	93	95
VIII	0	0	0	0	8	16	28	44	57	65	73	79	88	93	100	100	100
IX	0	0	0	0	8	18	28	42	60	71	78	85	93	100	100	100	100
X	0	0	0	0	0	3	9	17	24	29	36	46	55	67	82	100	100
XI	0	0	0	0	0	4	10	21	33	42	48	53	57	62	67	76	100
XII	0	0	0	0	3	7	18	30	41	48	56	64	73	84	100	100	100
XIII	0	0	0	0	0	0	5	11	16	22	28	38	50	71	100	100	100
XIV	0	0	0	0	0	2	10	19	28	37	46	54	64	74	85	93	100
XV	0	0	0	0	0	0	7	15	24	32	37	44	55	76	92	100	100
XVI	0	0	0	0	0	0	8	17	28	37	46	56	65	78	88	96	100
XVII	0	0	0	0	0	0	7	24	47	61	64	68	74	83	100	100	100
XVIII	0	0	0	0	0	0	6	11	18	26	36	47	61	82	100	100	100
XIX	0	0	0	0	0	0	2	5	8	14	20	31	43	55	68	80	96
XX	0	0	0	0	0	0	1	2	5	9	16	28	38	53	73	90	98

TABLE LI

DIFFERENCE BETWEEN 50% AND THE PER CENT PASSING AT EACH AGE

Designs (No.)	Age																
	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
I	−50	−50	−50	0	33	43	48	50	50	50	50	50	50	50	50	50	50
II	−50	−50	−50	−28	−1	15	28	39	44	46	47	48	49	50	50	50	50
III	−50	−50	−50	−31	−4	18	30	36	41	44	46	47	49	50	50	50	50
IV	−50	−50	−50	−44	−30	−16	0	18	29	35	38	43	47	50	50	50	50
V	−50	−50	−50	−50	−40	−25	−10	20	29	33	36	38	44	50	50	50	50
VI	−50	−50	−50	−46	−39	−30	−17	0	16	23	28	34	43	50	50	50	50
VII	−50	−50	−50	−50	−46	−42	−37	−30	−21	−12	−2	7	15	25	34	43	45
VIII	−50	−50	−50	−50	−42	−34	−22	−6	7	15	23	29	38	43	50	50	50
IX	−50	−50	−50	−50	−42	−32	−22	−8	10	21	28	35	43	50	50	50	50
X	−50	−50	−50	−50	−50	−47	−41	−33	−26	−21	−14	−4	5	17	32	50	50
XI	−50	−50	−50	−50	−50	−46	−40	−29	−17	−8	−2	3	7	12	17	26	50
XII	−50	−50	−50	−50	−47	−43	−32	−20	−9	−2	6	14	23	34	50	50	50
XIII	−50	−50	−50	−50	−50	−50	−45	−39	−34	−28	−22	−12	0	21	50	50	50
XIV	−50	−50	−50	−50	−50	−48	−40	−31	−22	−13	−4	4	14	24	35	43	50
XV	−50	−50	−50	−50	−50	−50	−43	−35	−26	−18	−13	−6	5	26	42	50	50
XVI	−50	−50	−50	−50	−50	−50	−42	−33	−22	−13	−4	6	15	28	38	46	50
XVII	−50	−50	−50	−50	−50	−50	−43	−26	−3	11	14	18	24	33	50	50	50
XVIII	−50	−50	−50	−50	−50	−50	−44	−39	−32	−24	−14	−3	11	32	50	50	50
XIX	−50	−50	−50	−50	−50	−50	−48	−45	−42	−36	−30	−19	−7	5	18	30	46
XX	−50	−50	−50	−50	−50	−50	−49	−48	−45	−41	−34	−21	−12	3	23	40	48

TABLE LII

P. E. EQUIVALENTS OF ITEMS IN TABLE LI

Designs (No.)	Age																
	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
I	?	?	4.6	0	−1.4	−2.2	−3.0	−4.6	−?	−?	−?	−?	−?	−?	−?	−?	−?
II	?	?	4.6	1.1	.0	− .6	−1.1	−1.8	−2.3	−2.6	−2.8	−3.0	−3.5	−4.6	−?	−?	−?
III	?	?	4.6	1.3	.1	− .7	−1.2	−1.6	−2.0	−2.3	−2.6	−2.8	−3.5	−4.6	−?	−?	−?
IV	?	?	4.6	2.3	1.2	.6	0	− .7	−1.2	−1.5	−1.7	−2.2	−2.8	−4.6	−?	−?	−?
V	?	?	?	4.6	1.9	1.0	.4	− .8	−1.2	−1.4	−1.6	−1.7	−2.3	−4.6	−?	−?	−?
VI	?	?	4.6	2.6	1.8	1.2	.7	0	− .6	− .9	−1.1	−1.5	−2.2	−4.6	−?	−?	−?
VII	?	?	?	4.6	2.6	2.1	1.7	1.2	.8	.5	.1	− .3	− .6	−1.0	−1.5	−2.2	−2.4
VIII	?	?	?	4.6	2.1	1.5	.9	.2	− .3	−6	− .9	−1.2	−1.7	−2.2	−4.6	−?	−?
IX	?	?	?	4.6	2.1	1.4	.9	.3	− .4	− .8	−1.1	−1.5	−2.2	−4.6	−?	−?	−?
X	?	?	?	?	4.6	2.8	2.0	1.4	1.0	.8	.5	.1	− .2	− .7	−1.4	−4.6	−?
XI	?	?	?	?	4.6	2.6	1.9	1.2	.7	.3	.1	− .1	− .3	− .5	− .7	−1.0	−4.6
XII	?	?	?	4.6	2.8	2.2	1.4	.8	.3	.1	− .2	− .5	− .9	−1.5	−4.6	−?	−?
XIII	?	?	?	?	?	4.6	2.4	1.8	1.5	1.1	.9	.5	0	− .8	−4.6	−?	−?
XIV	?	?	?	?	4.6	3.0	1.9	1.3	.9	.5	.1	− .1	− .5	−1.0	−1.5	−2.2	−4.6
XV	?	?	?	?	?	4.6	2.2	1.5	1.0	.7	.5	.2	− .2	−1.0	−2.1	−4.6	−?
XVI	?	?	?	?	?	4.6	2.1	1.4	.9	.5	.1	− .2	− .6	−1.1	−1.7	−2.6	−4.6
XVII	?	?	?	?	?	4.6	2.2	1.0	.1	− .4	− .5	− .7	−1.0	−1.4	−4.6	−?	−?
XVIII	?	?	?	?	?	4.6	2.3	1.8	1.4	1.0	.5	.1	− .4	−1.4	−4.6	−?	−?
XIX	?	?	?	?	?	4.6	3.0	2.4	2.1	1.6	1.2	.7	.3	− .2	− .7	−1.2	−2.6
XX[1]	?	?	?	?	?	4.6	3.5	3.0	2.4	2.0	1.5	.8	.5	− .1	− .9	−1.9	−3.0

TABLE LIII

P. E. INTERVALS BETWEEN CONSECUTIVE AGES

Design Number	Ages															
	3–4	4–5	5–6	6–7	7–8	8–9	9–10	10–11	11–12	12–13	13–14	14–15	15–16	16–17	17–18	18–19
I . .	?	?	4.6	1.4	.8	.8	1.6	?	?	?	?	?	?	?	?	?
II . .	?	?	3.5	1.1	.6	.5	.7	.5	.3	.2	.2	.5	1.1	?	?	?
III . .	?	?	3.3	1.2	.8	.5	.4	.4	.3	.3	.2	.7	1.1	?	?	?
IV . .	?	?	2.3	1.1	.6	.6	.7	.5	.3	.2	.5	.6	1.8	?	?	?
V . .	?	?	?	2.7	.9	.6	1.2	.4	.2	.2	.1	.6	2.3	?	?	?
VI . .	?	?	2.0	.8	.6	.5	.7	.6	.3	.2	.4	.7	2.4	?	?	?
VII . .	?	?	?	2.0	.5	.4	.5	.4	.3	.4	.4	.3	.4	.5	.7	.2
VIII . .	?	?	?	2.5	.6	.6	.7	.5	.3	.3	.3	.5	.5	2.4	?	?
IX . .	?	?	?	2.5	.7	.5	.6	.7	.4	.3	.4	.7	2.4	?	?	?
X . .	?	?	?	?	1.8	.8	.6	.4	.2	.3	.4	.3	.5	.7	3.2	?
XI . .	?	?	?	?	2.0	.7	.7	.5	.4	.2	.2	.2	.2	.2	.3	3.6
XII . .	?	?	?	1.8	.6	.8	.6	.5	.2	.3	.3	.4	.6	3.1	?	?
XIII . .	?	?	?	?	?	2.2	.6	.3	.4	.2	.4	.5	.8	3.8	?	?
XIV . .	?	?	?	?	1.6	1.1	.6	.4	.4	.4	.2	.4	.5	.5	.7	2.4
XV . .	?	?	?	?	?	2.4	.7	.5	.3	.2	.3	.4	.8	1.1	2.5	?
XVI . .	?	?	?	?	?	2.5	.7	.5	.4	.4	.3	.4	.5	.6	.9	2.0
XVII . .	?	?	?	?	?	2.4	1.2	.9	.5	.1	.2	.3	.4	3.2	?	?
XVIII . .	?	?	?	?	?	2.3	.5	.4	.4	.5	.4	.5	1.0	3.2	?	?
XIX . .	?	?	?	?	?	1.6	.6	.3	.5	.4	.5	.4	.5	.5	.5	1.4
XX . .	?	?	?	?	?	1.1	.5	.6	.4	.5	.7	.3	.6	.8	1.0	1.1
Average .	?	?	3.1	1.7	.9	1.1	.7	.5	.3	.3	.3	.5	1.0	1.6	1.2	1.8

TABLE LIV

DISTANCES ABOVE ZERO OF THE MEDIANS OF EACH AGE

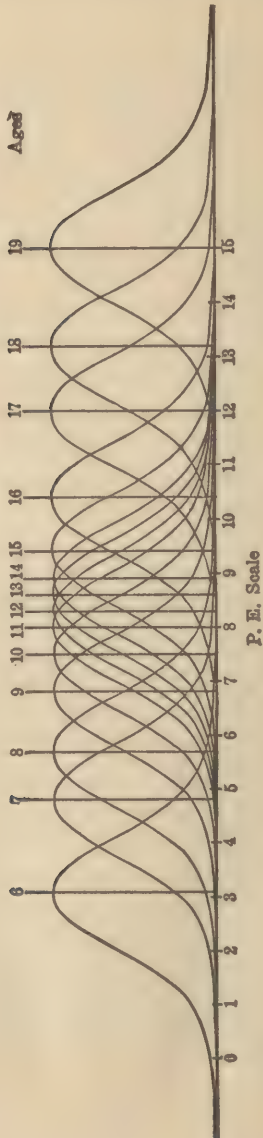
AGE	ABOVE ZERO	BELOW NEXT HIGHER AGE
4	?	?
5	0	3.1
6	3.1	1.7
7	4.8	.9
8	5.7	1.1
9	6.8	.7
10	7.5	.5
11	8.0	.3
12	8.3	.3
13	8.6	.3
14	8.9	.5
15	9.4	1.0
16	10.4	1.6
17	12.0	1.2
18	13.2	1.8
19	15.0	

The reader's attention is called to two matters of importance in the above standardization. First: Any age-group which manifested a difference between 50 per cent and the per cent passing at each age of plus or minus 50 per cent (see Table LI), was located on the P. E. Scale at 4.6 P. E. above or below the median for that grade. It is admitted that 4.6 P. E. is an arbitrary figure, but in this the writer has followed the custom of his predecessors. Second: The basic assumption underlying P. E. standardization has always been that distributions of mental ability follow the normal curve. Attention has already been called to the criticism launched by Boring against this hypothesis. We must admit a certain degree of error in our standards because of this peculiar inadequacy of our standardization technique, and yet we need not

GRAPH 40

Relation of Age

Distributions to Each Other



hesitate to use the results thus obtained, if we are at all times fully aware of their shortcomings. Until better instruments for norm-setting are placed at our disposal, we must utilize the best methods we now possess.

It may not be out of place here to call attention to a common fallacy, frequently observed in the writings and the methodology of present-day devisers of tests, which pertains particularly to the matter of "weighting." The situation is strikingly evident in group testing. For example, take one recently devised group-test for mental ability.[1] It is a booklet containing six exercises. Exercise One is a list of questions. The answers "Yes-No" are tabulated in a column at the right. The child performing the test is asked to read the questions carefully one at a time and draw a line under the right answer to each question. This question-exercise consists of *forty* items graded in order of difficulty. (Question 1 is, "Do dogs run?" — Question 40 is, "Are judicial decisions ever enforced?") Exercise Two, however, consists of only *twenty* arithmetic problems, also graded according to difficulty. (Example 1 is, "How many are 30 men and 7 men?" and example 20 is, "A submarine makes 8 miles an hour under water and 15 miles on the surface. How long will it take to cross a 100-mile channel, if it has to go two-fifths of the way under water?") A time limit of five minutes is allowed the pupil for each of these two exercises. The unfairness in the scoring lies in the fact that in adding up the total score each question successfully answered in Exercise One has equal value with each

[1] Haggerty, "Intelligence Examination," Delta 2, World Book Co., 1920.

question successfully answered in Exercise Two. It is true that no attempt has been made to weight these exercises. And the impression becomes fixed that no weighting is better than one even crudely constructed. This was the apparent decision of the committee which drafted the Army Tests.

The reason for mentioning this matter here is to justify the writer's procedure in attempting some sort of weighting process for each of the designs, and not only as regards successful performance of each of the designs, but as well for the time spent and the number of moves made on each design. It is the writer's contention that some rational and statistical basis for weighting is to be preferred to no weighting of any sort.

To come back then to our illustration. In the abovementioned group-test a child may answer twenty questions in Exercise One during the five minutes allowed, but in Exercise Two, the arithmetic problems, he may barely pass eleven. In summating the total score achieved, the answers to the questions inevitably have twice the importance as compared with the answers to the arithmetic problems. In other words, an item in the list of arithmetic problems has about half the weight of an item in the list of questions. In view of the lack of information on this particular matter, there seems to be no rational or statistical basis for this apparently arbitrary number of items in each exercise.

It is the writer's contention, therefore, that some reasonable attempt at weighting separate tests in a series is justifiable; that there is no escaping some sort of weighting; and that between the two evils of no

weighting at all, and one rationally and statistically arrived at, the latter certainly is to be preferred.[1]

In standardizing the block designs by the P. E. Scale Method, the writer has not followed exactly the orthodox procedure of Buckingham, Trabue, and Woody. If we keep in mind the fact that educational statistics pertaining to mental or pedagogical achievement are inclined to be far less accurate than physical or chemical statistics, it seems unnecessary to utilize extra refinements of standardization in order to obtain satisfactory and workable norms. It would, perhaps, be no exaggeration to state that we have erred as frequently in over-refining our statistical procedures as in not refining them enough.

Another matter worthy of note is that in the block-design tests the intervals between consecutive ages are *not* equal. The assumption frequently made is that increments of mental ability between ages *are* equal. The progress of the intervals for block-design ability is indicated in Graph 41. It appears that the differences are far more apparent between the ages 4 to 11 and 14 to 19 than between the ages 11 to 14. The elimination of a few designs in the middle range, from designs 5 to 15, might remedy this fault, if such it be.

The zero point on our P. E. Scale is very easily located from the tables of data. It is the age at and below which no children pass a given design.

[1] The writer does not wish to be understood as adversely criticizing the Haggerty Intelligence Examination. The tests, as such, are excellent, but I am convinced that even their author would admit that the norms and methods of scoring are open to improvement.

TABLE LV

LOCATION ABOVE O P. E. OF EACH DESIGN AT EACH AGE

DESIGN (No.)	AGE																FINAL P. E. VALUE
	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
I			3.1	3.4	3.5	3.8											3.45
II			4.2	4.8	5.1	5.7	5.7	5.7	5.7	5.8	5.9	5.9					5.45
III			4.4	4.9	5.0	5.6	5.9	6.0	6.0	6.0	6.1	5.9					5.58
IV			5.4	6.0	6.3	6.8	6.8	6.8	6.8	6.9	6.7	6.6					6.51
V				6.7	6.7	7.2	6.7	6.8	6.9	7.0	7.2	7.1					6.92
VI			5.7	6.6	6.9	7.5	7.5	7.4	7.4	7.5	7.4	7.2					7.11
VII				7.4	7.8	8.5	8.7	8.8	8.8	8.7	8.6	8.8	9.4	10.5	11.0	12.6	9.20
VIII				6.9	7.2	7.7	7.7	7.7	7.7	7.7	7.7	7.7	8.2				7.62
IX				6.9	7.1	7.7	7.8	7.6	7.5	7.5	7.4	7.2					7.41
X					8.5	8.8	8.9	9.0	9.1	9.1	9.0	9.2	9.7	10.6			9.19
XI					8.3	8.7	8.7	8.7	8.6	8.7	8.8	9.1	9.9	11.3	12.2		9.36
XII				7.6	7.9	8.2	8.3	8.3	8.4	8.4	8.4	8.5	8.9				8.29
XIII						9.2	9.3	9.5	9.4	9.5	9.4	9.4	9.6				9.41
XIV					8.7	8.7	8.8	8.9	8.8	8.7	8.8	8.9	9.4	10.5	11.0		9.20
XV						9.0	9.0	9.0	9.0	9.1	9.1	9.2	9.4	9.9			9.18
XVI						8.9	8.9	8.9	8.8	8.7	8.7	8.8	9.3	10.3	10.6		9.19
XVII						9.0	8.5	8.1	7.9	8.1	8.2	8.4	9.0				8.40
XVIII						9.1	9.3	9.4	9.3	9.1	9.0	9.0	9.0				9.15
XIX						9.8	9.9	10.1	9.9	9.8	9.6	9.7	10.2	11.3	12.0	12.4	10.42
XX						10.3	10.5	10.4	10.3	10.1	9.7	9.9	10.3	11.1	11.3	12.0	10.53

GRAPH 41

Location of Age Medians on

P. E. Scale

(Showing Size Relation of Intervals)

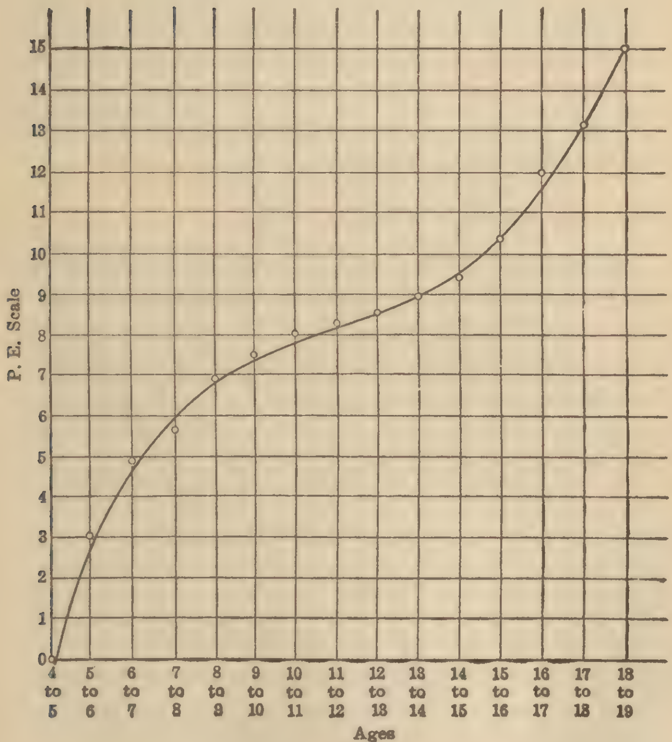


Table LV may then be constructed from the data of Tables LII and LIV, giving the location above zero P. E. of each design at each age. In the last column to the right (Table LV) are given the final P. E. values of each design, obtained by averaging the findings for that design at the various age levels. We differ, again, at this place in our procedure from that followed by Thorndike's students, by not rejecting any of our individual P. E. values, and by giving each value equal weight.

Graph 39 is a pictorial arrangement of the twenty designs on the P. E. Scale.

(*b*) *Time and Move Norms*

As has already been stated, the writer felt in the very early stages of the standardization process, that time and number of moves were very important diagnostic criteria which could not conscientiously be left out of consideration in the final assignment of a score for a specific performance. Hoping for some clue whereby the three criteria: successful performance, time, and moves, could be amalgamated satisfactorily in a single score-value, the record of every test made contained a complete statement of these three reactions for each design. It was difficult in the preliminary stages of the investigation to state dogmatically which of the three criteria was most and which least important. But at the conclusion of our experimentation sufficient objective evidence had been accumulated to answer these questions more or less adequately.

One clue which seemed more worth while developing than any other was the relation between the above-

mentioned three criteria and life age. (Mental age would, undoubtedly, have served as well.) If success, time, and moves were diagnostic tools of varying importance, it seemed plausible that these differences would become apparent if correlated either with life age or with mental age.

This was therefore done with the following somewhat startling results:

CORRELATIONS

		r	*P.E.*
Life Age and Raw Scores	(241 Cases Public School Children, "at grade")	+.71,	±.02
Life Age and Time	(Averaged first the individual coefficients and then the individual P. E.'s for each of the twenty designs. *Note:* Only those coefficients were used where *r* was at least four times the P. E. The results for five designs were therefore not used. See Table LVII.)	+.38,	±.06
Life Age and Moves	(Averaged the individual coefficients for *r* and P. E. as above. *Note:* Only those coefficients were used where *r* was at least twice as large as the P. E. The results for six designs were therefore not used. See Table LVII.)	+.20,	±.07
Binet Mental Age and Raw Srores	(366 cases. Included are all the subjects of this study.)	+.82,	±.01

From these figures it was possible to conclude for this test that success, as such, is about twice as important a diagnostic criterion as time (ratio 71 to 38), and that time is about twice as important as moves (ratio 38 to 20). In other words, giving moves a value or a weight

of 1, time should, according to the above ratio, receive a weight of 2, and success a weight of 4. It is this relationship that is referred to when the ratios 4:2:1 are stated in various sections of this monograph.

Scientific attitude precludes the existence of bias. Nevertheless, it is the rare experimenter who can count himself free from all sorts of fixed notions. The writer's conception of the relative importance of time and moves squared with the prevalent opinion that moves are more important diagnostic-indicators of intelligence than time. The poor correlations between "speed" tests versus "power" tests helped confirm this belief. Besides, it had come to be accepted that to attempt the combination of such utterly diverse entities as success, time, and moves was nothing short of quixotic.

In presenting the above figures there has been no desire to enter into a controversy with common opinion. The facts are presented as facts, and as such should receive their requisite evaluation. The writer knows of no other instance where an attempt has been made by similar impersonal methods to determine objectively the weights to be accorded the three intelligence-criteria mentioned above, and will await with interest the efforts of others in this direction.

The correlation between life age and the raw scores (the factors of time and moves not having been taken into consideration) is presented in Table LVI. (All of the correlations in this monograph have been determined by means of the formula: $r = \frac{\Sigma xy}{\sqrt{\Sigma x^2} \cdot \sqrt{\Sigma y^2}}$, modified by the writer. See Appendix II.)

TABLE LVI

LIFE AGE AND RAW SCORES CORRELATION

Designs Completed	Life Age												Total
	6	7	8	9	10	11	12	13	14	15	16	17	
0 . . .	8	3	1	1									13
1 . . .	2	8	4		3	2		1	1				21
2 . . .	3	3	5	7	1								19
3 . . .	2	7	5	2	1		1	1		1			20
4 . . .		1	2	3	2		1	1	1	1			12
5 . . .		2	3	4	4	2							15
6 . . .		1	3	2	3	4	2	1		2			18
7 . . .		1	2	1	3	2	3	1	3				16
8 . . .			2		1	1	1	2		1			8
9 . . .			1	1	2	1		1	2	1	1		10
10 . . .						2	1	4		1	1		9
11 . . .				1	3	2	1	1					8
12 . . .					1	1	2	1	1		1		7
13 . . .				1		3	2		1		1		8
14 . . .					1	2	2	2					7
15 . . .							1	2	5		1		9
16 . . .					1	2	3		1				7
17 . . .				1	1	1	1	1	1	2	1		9
18 . . .								2	3	2	2		9
19 . . .					1		1	1	2	1	1	1	8
20 . . .									2	3	2	1	8
Total .	15	26	28	24	28	25	22	22	23	15	11	2	241

$r = +.71$ P. E. $= \pm .02$

The 241 children included in Table LVI were only those of the public school group who were located at the proper grade for their life age.

The correlations between life age and time and between life age and moves for each of the twenty designs is presented in Table LVII, together with their respective probable errors. All the correlations are positive, except where otherwise indicated.

In the fourth and eighth columns of the table are presented the ratios between the coefficients for time and

TABLE LVII

CORRELATIONS

Life Age with Time and with Moves

Design Number	Correlation Age-Time	Correlation Age-Moves	Ratio Time: Moves	Design Number	Correlation Age-Time	Correlation Age-Moves	Ratio Time: Moves
	r — P. E.	r — P. E.			r — P. E.	r — P. E.	
I . .	.38 ± .04	.22 ± .04	1.7 : 1	XI .	.46 ± .06	.18 ± .08	2.6 : 1
II . .	.45 ± .04	.26 ± .04	1.7 : 1	XII .	.41 ± .06	.27 ± .07	1.5 : 1
III . .	.38 ± .04	.17 ± .05	2.2 : 1	XIII .	.59 ± .06	.23 ± .09	2.6 : 1
IV . .	.53 ± .04	.35 ± .05	1.5 : 1	XIV .	.45 ± .07	.21 ± .08	2.0 : 1
V . .	.50 ± .04	.27 ± .05	1.9 : 1	XV .	.46 ± .07	.22 ± .08	2.1 : 1
VI . .	.34 ± .05	.12 ± .06	2.8 : 1	XVI .	.20 ± .08	.03 ± .08	
VII . .	.31 ± .07	.11 ± .07		XVII .	.34 ± .05	.45 ± .06	1.2 : 1
VIII . .	.16 ± .06	.00 ± .06		XVIII .	.10 ± .09	.02 ± .09	
IX . .	.47 ± .05	.31 ± .05	1.5 : 1	XIX .	.41 ± .10	.49 ± .09	0.8 : 1
X . .	.31 ± .08	.08 ± .08		XX .	.14 ± .13	−.10 ± .13	
				Average	.38 ± .06	.20 ± .07	1.9 to 1

moves for each design. These ratios are finally averaged, with the result that time appears a better diagnostic criterion than moves to the extent of 2 to 1. In averaging the coefficients for life age and time the results for Designs VIII, X, XVI, XVIII and XX were not included, because their coefficients were less than four times their P. E.'s. And in averaging the coefficients for life age and moves, the results for Designs VII, VIII, X, XVIII and XX were not included, because their coefficients were less than twice their respective P. E.'s. Consequently in averaging the final ratios for time and moves the ratios for Designs VII, VIII, X, XVI, XVIII and XX were not used.

No other methods have been utilized to check the validity of this procedure for determining the relative weights to be apportioned success, time, and moves. However, the method outlined above is objective and free from personal bias.

As regards sex differences for time and moves, there were no greater differences apparent than might be expected from the expected errors of sampling.

To go on then to the separate norms for time and for moves, in Table LVIII are found the figures for the average times (arithmetic mean), and in Table LIX the average moves for each design at each chronological age, from 6 to 17. Thus for Design I, at age 6, 72 seconds are taken for successful completion, as compared with 22 seconds at age 11, and with only 7 seconds at age 17. And again, for Design XIX, at age 9, 258 seconds (over 4 minutes) are necessary for successful completion, as compared with 171 seconds at age 14, and only 99 seconds

TABLE LVIII

TIME NORMS (SECONDS)

LIFE AGE	Designs I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI	XVII	XVIII	XIX	XX
6	72	73	59																	
7	34	72	70	79	92	122	108	76	77			203								
8	34	48	44	74	91	79	92	74	97	142	163	232								
9	24	38	32	77	69	48	31	33	89	123	147	154	132	219	196	141	177	212	258	
10	24	32	43	62	57	53	96	46	71	154	114	137	144	202	135	157	168	203	192	
11	22	36	25	39	69	49	64	61	55	113	116	126	163	159	202	170	166	171	117	215
12	24	27	25	42	48	52	70	45	55	125	115	142	181	169	149	175	142	162	185	171
13	18	21	23	34	40	49	81	47	43	144	116	131	121	137	161	170	116	174	197	189
14	15	26	19	28	30	24	66	47	43	116	85	93	111	120	142	138	111	181	171	131
15	15	18	26	37	36	54	60	46	30	100	90	99	67	142	126	137	106	168	151	200
16	10	18	14	23	29	29	52	31	30	106	92	99	79	144	148	154	118	186	148	164
17	7	15	9	12	11	16	65	43	14	82	43	55	44	122	105	112	78	141	99	141

TABLE LIX

Move Norms (Moves)

Life Age	Designs																			
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI	XVII	XVIII	XIX	XX
6	8.9	11.7	8.3																	
7	5.4	9.6	9.7	10.4	15.0	18.0	17.0	17.5	17.3			22.0								
8	5.6	6.9	7.8	12.5	13.7	13.5	14.2	10.4	14.0	19.0	27.0	21.0								
9	5.0	5.6	6.2	12.2	11.9	11.0	5.0	6.0	15.6	19.0	19.5	18.3	15.0	35.3	32.0	22.0	32.5	34.5	42.0	
10	5.6	6.2	9.3	11.7	10.4	12.4	18.6	8.1	12.4	19.4	16.7	17.0	16.3	34.8	30.2	32.0	30.0	33.0	30.0	
11	4.7	7.0	6.5	8.5	12.9	12.5	12.7	10.9	9.7	24.2	18.9	18.1	26.3	29.7	32.2	34.0	35.1	29.0	30.0	23.0
12	5.6	6.2	7.7	8.7	9.9	10.8	13.4	8.1	9.0	20.9	18.9	19.3	26.8	32.1	26.0	34.7	30.5	30.4	32.2	27.5
13	4.9	5.7	6.0	7.7	9.0	12.2	15.6	8.4	7.7	24.1	18.4	19.3	17.2	27.0	30.8	32.3	27.4	28.6	33.0	34.3
14	4.6	6.0	6.2	8.6	8.3	8.1	13.3	9.0	9.7	22.4	17.8	14.7	16.0	27.7	30.5	29.5	26.5	32.6	28.9	22.9
15	4.6	4.8	8.1	8.3	9.8	13.6	13.4	7.7	7.1	18.8	17.0	16.6	13.8	31.4	27.0	29.7	24.7	28.3	27.6	34.8
16	4.3	5.6	5.3	6.7	9.0	8.9	12.2	6.2	7.2	20.0	16.2	14.3	13.9	29.4	28.7	33.3	24.4	37.3	25.5	28.6
17	4.0	5.0	5.0	7.5	4.5	7.0	9.0	11.5	4.5	19.0	10.0	11.5	10.0	31.5	22.0	27.0	20.0	29.5	21.0	28.5

at age 17. So, too, for moves. For Design I, at age 6, 8.9 moves are made before success is finally accomplished, as compared with 5.6 moves at age 10, and with 4.6 moves at age 14. And again, for Design XIX, at age 9, 42.0 moves (almost 3 moves to each block) are made before the design is successfully completed, as compared with 28.9 moves at age 14, and only 21 moves at age 17.

An analysis of sex differences reveals the following:

For designs I–X, ages 6–8, the tendency is greater for boys to give longer times on the same designs, although the frequency for giving long reaction times is about the same for both groups. For designs I–X, ages 9–13, boys more frequently exceed the times of the girls (ratio: boys, 31; girls, 18), but the amount of excess is about the same (ratio: boys, 15″; girls, 18″). For designs I–X, ages 14–16, and for designs XI–XX, ages 9–13, no tendency is evident for boys to exceed girls in time per design, or vice versa, the excess being about the same. And in designs XI–XX, ages 14–16, boys more frequently give longer times (ratio: boys, 24; girls, 6), but the girls give longer times whenever they exceed in time (ratio: boys, 27″; girls, 34″). In general, it may be stated that boys tend to give longer reaction times more frequently than girls, but when girls take longer the excess in time is greater than the excess in time for boys.

As regards sex differences in moves, there are no appreciable differences except for designs I–IX at ages 6–8, and 9–13. In the first instance, the tendency to make a greater number of moves is about the same, but the excess is greater for boys than for girls; and in the latter

case, the tendency to make a greater number of moves is stronger for boys than for girls, but the excess of moves is about the same.

Tables LX, LXI, and LXII give the correlations between time and moves for the four-block, nine-block, and sixteen-block designs. The correlations are high, as one would expect.

TABLE LX

CORRELATION BETWEEN TIME AND MOVES

(4-block designs)

Moves	Time					Total
	1″–30″	31″–1′	1′1″–1′30″	1′31″–2′	2′1″–2′30″	
4 . . .	140	10		1		151
5– 8 . . .	192	90	8	3	1	294
9–12 . . .	13	10	27	6	4	60
13–16 . . .		3	18	11	7	39
17–20 . . .			10	8	6	24
21–24 . . .			6	2	7	15
25–28 . . .			1		1	2
29–32 . . .			1	1	2	4
33 and over .					1	1
Total .	345	113	71	32	29	590

$r = + .76, \pm .01$

Four groups of workers may be segregated from these figures by horizontal and vertical divisions of the tables. One group which works rapidly and accurately, and consists almost entirely of those who succeed. A second group which works somewhat too rapidly perhaps, and attempts numerous varieties of block-combination. A third group which works slowly, makes few moves and appears deliberative in procedure. And a fourth group which not only works slowly, but also attempts numerous combinations, many bordering on the irrational.

TABLE LXI

Correlation between Time and Moves

(9-block designs)

Moves	Time								Total
	1"-30"	31"-1'	1'1"-1'30"	1'31"-2'	2'1"-2'30"	2'31"-3'	3'1"-3'30"	3'31"-4'	
9		4	1						5
10–15	1	4	9	9	2	2			27
16–21		2	10	12	12	7			43
22–27			1	3	3	6	4	2	19
28–33					1	4	1	1	7
34–39						1			1
40–45								1	1
Total	1	10	21	24	18	20	5	4	103

$r = +.58, \pm .04$

TABLE LXII

Correlation between Time and Moves

(16-block designs)

Moves	Time								Total
	31"-1'	1'1"-1'30"	1'31"-2'	2'1"-2'30"	2'31"-3'	3'1"-3'30"	3'31"-4'	4'1"-4'30"	
17–20		3	1	1	1				6
21–24	1	13	10	6	2				32
25–28		1	11	9	11		2		34
29–32			2	7	6	7	1		23
33–36				3	8	8	4		23
37–40				2	5	2	3	1	13
41–44					2	3	2	1	8
45–48					1		3	1	5
49–52									
53–56							1	1	2
Total	1	17	24	28	36	20	16	4	146

$r = +.70, \pm .03$

(c) *Result: The P. E. Scale*

On the basis of these results bearing upon the relation between time and moves, and in view of the earlier findings, the following score-card was devised (Table LXIII). The attempt was made to apportion due weight to (a) score value, (b) time, and (c) moves in the ratio of about 4 : 2 : 1.

TABLE LXIII

Score Card

Design Number	P.E. Score Value	Excess Time: Deduct 1 Point	Excess Time: Deduct 2 Points	Excess Moves Deduct 1 Point
I	3	21″ and over		6 and over
II	5	31″ and over		7 and over
III	6	21″– 35″	36″ and over	8 and over
IV	7	31″–1′	1′ 1″ and over	10 and over
V	7	36″–1′ 5″	1′ 6″ and over	11 and over
VI	7	36″–1′	1′ 1″ and over	12 and over
VII	9	56″–1′10″	1′11″ and over	15 and over
VIII	8	41″– 55″	56″ and over	10 and over
IX	7	41″–1′10″	1′11″ and over	11 and over
X	9	1′56″–2′10″	2′11″ and over	22 and over
XI	9	1′36″–2′	2′ 1″ and over	18 and over
XII	8	1′46″–2′30″	2′31″ and over	19 and over
XIII	9	1′31″–2′30″	2′31″ and over	18 and over
XIV	9	2′26″–2′40″	2′41″ and over	32 and over
XV	9	2′26″–2′40″	2′41″ and over	30 and over
XVI	9	2′21″–2′33″	2′34″ and over	31 and over
XVII	8	1′56″–2′25″	2′26″ and over	30 and over
XVIII	9	2′41″–3′ 0″	3′ 1″ and over	32 and over
XIX	10	2′41″–3′ 5″	3′ 6″ and over	31 and over
XX	11	2′41″–2′55″	2′56″ and over	30 and over

Maximum Score = 159 Points

The information in the last three columns of the preceding table was obtained by plotting first the norm data for each design at each age, both for time and for moves. All the curves were then smoothed. After that the year-range of the curve was divided into as nearly

three equal year-divisions as possible. Thus for design XI the total year-range for the time-and-life-age curve was nine years (age 8 to age 17). All times, then, falling within the range of years between 14 and 17 were allowed as legitimate. One point of the score value of the design, however, was to be deducted if the time taken fell within the range of years from 11 to 14, and two points were to be deducted if the time taken fell within the earliest range of years, namely, between the years 8 and 11. The same procedure was utilized for determining the deductions for excess time on the other designs, except designs I and II for which the method pursued for moves was followed. For the one point deduction for excess moves the year-range was divided into two parts. The range of moves covering the higher half of years was regarded as legitimate, whereas the range of moves covered by the lower half of years was regarded in excess and one point was designated as a reasonable deduction from the score value of that design.

Other methods for arriving at an objective criterion for making subtractions for excess time and moves were possible.[1] For example, instead of the *year-range* being divided into three equal divisions, the *score-range* might have been thus divided. Another method possible would have been to take the score-range and divide it into four equal parts. Two points would then be de-

[1] It did not seem that any other possible methods possessed any superior value over the one actually utilized, and in any event, whatever the procedure for penalizing excess moves and time, the chances are very great that no great differences would have resulted in the ultimate score assignments to each performance of the block-design tests, and in the final standards.

ducted for the first quarter of the range falling at the lowest ages, one point would then be deducted from the middle range of 50 per cent, and all times falling within the upper range of 25 per cent would have been regarded as legitimate.

Utilizing the scoring instrument thus devised, all the tests were then re-scored. The original correlation between the Binet Scale and the number of designs successfully completed (already designated as the "raw score," time and moves not being taken into account), was + .82, ± .01. The table is presented herewith:

TABLE LXIV

CORRELATION BETWEEN BINET MENTAL AGE AND RAW BLOCK-DESIGN SCORES

Binet Age

RAW SCORES	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	TOTAL
0	3	7	14	14	3	1		1										43
1				7	9	9	2	2	1	1	1							32
2				5	5	7	7	3	3									30
3				4	6	6	5	1	1		2	1						26
4					1	7	6	5	2	1	1	1						24
5				1	3	9	6	5	1	1	1							27
6					1	5	4	4	4	3	2		1					24
7						4		6	6	3	1		1					21
8						2	3	3	1	1				1				11
9					1	1		2	4	1	1	2						12
10								1	3	3	3		1					11
11								4	4	1		1						10
12							1	1	2	1		1	1	1				8
13							1		4	2	2	1						10
14							2	4	1	5	1	1						14
15								1	1	4	1	2		1				10
16									1	2		4	1	1				9
17								1	3	1	2	1	1	1			1	11
18										1	1		3	1	3	2		11
19									2			2	1	2	1	1	2	11
20											1		1	3	3	1	2	11
Total . .	3	7	14	31	29	51	37	44	44	31	20	17	11	11	7	4	5	366

Although some of the values in the preceding table are scattered, there is a marked consistency in the increase of raw score with increase of mental age.

Upon rescoring the 366 tests in order to assign due weight to success, time, and moves, the coefficient of correlation is not changed to any appreciable amount, but there is a marked decrease in the fluctuation, or spread, of scores over any given range of years. This is more apparent after the elimination of certain designs, discussed more in detail in the succeeding section of the monograph. The correlation table follows:

TABLE LXV

CORRELATION BETWEEN BINET MENTAL AGE AND SCORE POINTS

Score Points	Mental Age																	Total
	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
0	3	7	14	14	3	1		1										43
1–10 . . .				15	19	20	12	6	5	1	1	1						80
11– 20 . .				2	3	13	11	9	2	2	4	1						47
21– 30 . .					3	9	7	7	5	4	3		2					40
31– 40 . .						7	2	4	6	3				1				23
41– 50 . .					1	1	1	4	5	1	1	1						15
51– 60 . .								4	6	4	3	1	1					19
61– 70 . .							2	3	2	1	1	1	1	1				12
71– 80 . .									4	3	1	1						9
81– 90 . .							2	2	2	3	1	2						12
91–100 . .								3	2	5		3	1	2				16
101–110 . .									1	3	2	2						8
111–120 . .									1	1	1	2	1	2				8
121–130 . .								1	3		1		1		1		2	9
131–140 . .												2	2	1	2	2	2	11
141–150 . .											1		2	1	2	1		7
151–160 . .														3	2	1	1	7
Total . .	3	7	14	31	29	51	37	44	44	31	20	17	11	11	7	4	5	366

$r = +.83, \pm .01$

In order to appreciate the significance of this coefficient, the following list of correlations between the Stanford-Binet test and other intelligence tests is presented:

Block-design Test and Binet Test . . + .83
Army Alpha and Binet Test,[1] . + .80 to + .90
Army Beta and Binet Test,[1] . . . + .73
Stanford-Binet and Stanford-Binet (repeated tests),[1] + .94 to + .97
Stanford-Binet Entire and Stanford-Binet Abbreviated (2 tests per year)[1] + .92
Each Army Performance Test and Stanford-Binet[1]:

1. Ship Test	
2. Manikin and Feature Profile Test	
3. Cube Imitation Test	Lowest Correlation, +.48
4. Cube Construction Test	
5. Form Board Test	
6. Reproducing Designs Test	
7. Digit-Symbol Test	
8. Maze Test	Highest Correlation, +.78
9. Picture Arrangement Test	
10. Picture Completion Test	

Abbreviated Army Performance Test (5 Tests) and Stanford Binet,[1] + .84
Stanford-Binet and Seguin Form Board,[2] . . + .55
Porteus and Stanford-Binet[2] { Boys . . . + .21; Girls . . . + .60 }
Healy's 3 Form Boards (Mare and Foal, Construction Test A, and Construction Test B) and Stanford-Binet Mental Age,[3] . + .45
Porteus Mental Age and Stanford-Binet Mental Age[3] . . . + .79

Army Alpha	and Ferguson Form Boards[4]	+ .51
Teachers Estimate		+ .50
Class Standing		+ .56

Grade Location and Form Board Ability,[4] + .80

[1] C. S. Yoakum and R. M. Yerkes: "Army Mental Tests" (New York, Holt, 1920), p. 20
[2] S. D. Porteus: "Porteus Tests — The Vineland Revision" (Vineland, 1919), p. 28, 19, 20.
[3] Elizabeth L. S. Ross: "Vocational Tests for Mental Defectives: Studies in Mental Inefficiency" (1921), 2: 1–6.
[4] G. O. Ferguson, Jr.: "A Series of Form Boards," *J. of Exp. Psychol.*, 1920, 3: p. 57.

TABLE LXVI

INTER-TEST CORRELATIONS [1]

(Table Rearranged)

	ARMY	THURSTONE	PRESSEY	INDIANA SURVEY	OTIS
Army		+ .60	+ .36	+ .36	+ .57
Thurstone	+ .60		+ .25	+ .25	+ .46
Pressey	+ .36	+ .25		+ .22	+ .44
Indiana Survey	+ .36	+ .25	+ .22		+ .34
Otis	+ .57	+ .46	+ .44	+ .34	

Pintner's Non-Language Group-Test and Stanford-Binet I. Q.,[2] + .66.

A cursory survey of the stragglers in the upper right-hand corner of the table (LXV) reveals children who obtain high Binet scores, but do poorly in these tests. These are generally children who are retarded mentally or are feeble-minded. There seems to be evidence that children of inferior endowment do poorer work with these tests than children of superior endowment *but of the same mental age.* So, too, the stragglers at the lower left-hand corner of the table are children who are foreign born, the language-handicapped, such as the Portuguese, Italians, Japanese, and Chinese, who do poorly with the Binet, yet obtain good scores on the designs.

The following are the median scores at each mental age. These medians were determined from a table of

[1] J. A. Clement: "Use of Mental Tests as a Supplementary Method for Making School Adjustments in Colleges." *Educ. Administr. and Superv.*, 1920, 6: p. 441.

[2] R. Pintner: "A Non-Language Group Intelligence Scale." *J. of Applied Psychol.*, 1919, 3: 199–214, p. 214.

original data, no grouping of scores beyond the one unit being made for scores 1 to 15. Beginning with score 16, however, the frequency intervals were 5 score-points. The median values in Table LXVII are, therefore, interpolated only from age ten up. The left half of the table presents actual values, and the right-hand portion the medians obtained from the smooth curve of Graph 42.

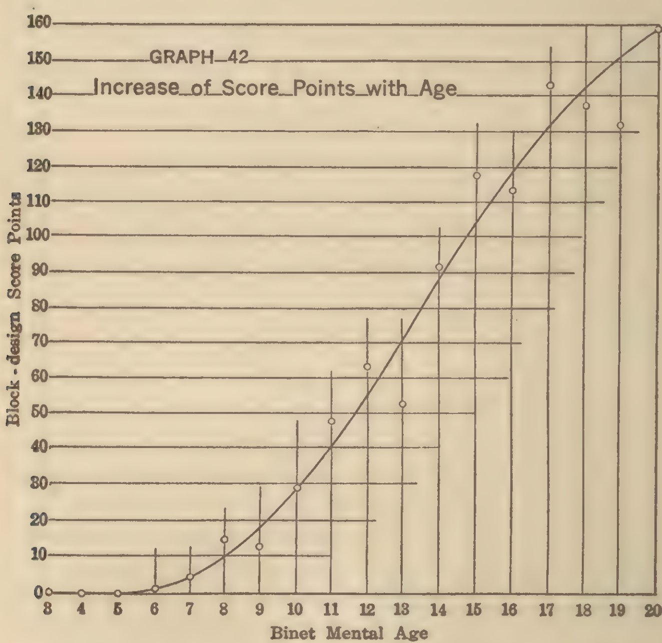
TABLE LXVII

MEDIAN SCORES AT EACH MENTAL AGE

Actual Medians		Smooth Curve Medians	
Age	Score Points	Age	Score Points
3, 4, 5	0	3, 4, 5	0
6	1	6	1
7	5	7	5
8	15	8	9
9	13	9	17
10	29	10	29
11	48	11	41
12	63	12	55
13	53	13	71
14	92	14	88
15	118	15	104
16	114	16	118
17	143	17	133
18	138	18	143
19	132	19	152
20	159	20	159

Graph 42 is the curve showing the increase of score with mental age. The dots represent the median scores at each age. The deviations between the curve and the actual medians are explained largely on the basis of idiosyncracies in the age-groups tested. Thus at age thirteen, the difference is quite striking. But by referring

back to the table showing the correlation between Binet age and life age (Table XIII) it will be observed that at thirteen the median mental age of the group examined (33 — 16 boys and 17 girls) was about twelve years, a



retardation of approximately one year. Referring back to the curve (Graph 42) for the performance of the 12-year-mental-age group, it will be observed that this group attains scores superior, on the average, to the 12-year standard, but inferior to the 13-year standard. The explanation is practically the same for the deviations at the other ages.

There was a twofold motive in further work to make the block designs a reliable and a serviceable test: In the first place, it was deemed wise to reduce the time for giving the test as much as possible without a sacrifice of reliability; and, on the other hand, there were a few designs which, if eliminated, would tend to decrease the scatter of scores over a given range of mental ages. It was for these reasons that further analyses were made, in order to determine which designs could be sacrificed in order to bring the above desirable conditions to pass. Space prohibits an extended discussion of the technique of elimination, but the main features of the revision are presented in the following section.

4. FINAL REVISION

(a) *Eliminations*

In reducing the number of designs it was recognized that the most desirable procedure to follow would be that which would result in the greatest reduction of scores for those of progressively poorer mental capacities, without affecting the scores of those of higher mental development. A general leveling of scores, without a compensating increase in the diagnostic importance of a unit score, as a rule, decreases the actual diagnostic value of any measuring scheme. The designs, therefore, which were considered for elimination were those falling within the middle range, namely, designs VII, XI, XIII, XIV, XV, XVI, and XVII. On the basis of various criteria which were utilized, those designs could more easily be sacrificed than the others.

Three elimination schemes were tried:

(*a*) the elimination of designs VII, XI, XIII, XVI, and XVII;
(*b*) the elimination of designs XI, XIII, XIV, XV, and XVII; and
(*c*) the elimination of designs XI, XIII, and XVII.

The final eliminations to be effected were to be those which would (*a*) most effectively reduce the scatter or range of scores from ages 3 to 19, (*b*) that would save time with least sacrifice of diagnostic efficiency, (*c*) that would result in a higher coefficient of correlation between block-design score and Binet mental age, and (*d*) one that would show an increased reliability of final score achieved.

Elimination Method (*a*). Thus elimination method (*a*) permitted a maximum score attainment of 115. The correlation beween the number of points lost (from the original score using the twenty designs) and Binet age was + .63.[1] The number of cases in which the score was affected was 140 or 38 per cent. Analyzing the results of this re-scoring method more closely, it becomes apparent that taking each mental age there has been a greater ratio of reduction in the scores of the lower ages (73 per cent) and a smaller ratio of reduction in the scores of the upper ages (27 per cent) — although the *absolute* values show a gradual increase in points lost with increase of age. The changed scores make for a

[1] This coefficient may also be interpreted as the correlation between these five designs taken alone and Binet mental age. This is a fair criterion of the reliability of the block-design tests as a whole, if these five designs are the poorest.

greater differentiation between the lower and the higher mental ages. Although the extremes of score were from 5 to 159, the latter score being 31 times the former, the extremes of per cent lost were from 8 to 41, only 5 times the former instead of 31.

Elimination Method (*b*). Rescoring the tests on the basis of the eliminations by method (*b*) the maximum score attainable is 115, the same as by method (*a*). Here again the correlation between block-design scores and Binet mental age was $+ .85 \pm .01$. In this instance, however, the scores at the lower ages were affected to the same extent as by method (*a*), but the scores at the higher ages were reduced in greater proportion than was the case with method (*a*), largely due to the limitation of designs eliminated to those above design X. Comparing the two methods, it was found that in 46 per cent of the cases, method (*a*) yielded higher scores than method (*b*), in 24 per cent of the cases method (*b*) yielded higher scores than method (*a*), and in 30 per cent of the cases, the scores by both methods were exactly the same.

Elimination Method (*c*). Based on the results of methods (*a*) and (*b*), method (*c*) is a compromise, only those designs being cast for elimination which were common to methods (*a*) and (*b*). By this method the maximum score attainable is 133, 26 less than when all the twenty designs are used. The correlation between block-design score points and Binet mental age is $+ .84 \pm .01$. A survey of the correlation table (Table LXVIII) reveals a somewhat greater consolidation, or a reduction in the scatter, of the tabulated scores about the mean trend as compared with the original correlation table

(Table LXV) which presents the scores based upon the use of the total twenty designs.

TABLE LXVIII

CORRELATION BETWEEN BINET MENTAL AGE AND SCORE POINTS
(Elimination Method (*c*))

Binet Mental Age

Score Points	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	Total
0 . . .	3	7	14	14	3	1		1										43
1– 10 . .				15	19	20	12	6	5	1	1	1						80
11– 20 . .				2	3	13	11	9	2	2	4	1						47
21– 30 . .					3	10	7	9	5	4	3		2					43
31– 40 . .						6	2	3	7	4				1				23
41– 50 . .					1	1	1	6	9	3	3	2						26
51– 60 . .							1	4	5	3	1	1	1					16
61– 70 . .							1	1	3	4	3	1	1	1				15
71– 80 . .							2	2	2	5		3		2				16
81– 90 . .								1		3	2	4	1					11
91–100 . .								1	3	2	1	1		2				10
101–110 . .								1	2		1	2	3		3	3	3	18
111–120 . .									1			1	2	1	1		1	7
121–130 . .											1		1	3	2		1	8
131 and over . .														1	1	1		3
Total . .	3	7	14	31	29	51	37	44	44	31	20	17	11	11	7	4	5	366

$r = +.84 \pm .01$

A comparison of the mental ratings resulting from this new array of designs with those in which all the designs were included, and with those in which eliminations (*a*) and (*b*) were made, revealed that the original determination of reducing the scatter of scores, of saving time without a sacrifice of diagnostic efficiency, of increasing the validity of the final score achieved, and of increasing the correlation between block-design scores and Binet mental age, had, to a degree been accom-

plished. The superiority of elimination method (*c*) is again made manifest from the following table, which shows what per cent the quartile deviation is of the maximum score (the score-range) at each age, for each of the elimination methods.

TABLE LXIX

PER CENT THAT Q. D. IS OF SCORE RANGE

	Age													
	6	7	8	9	10	11	12	13	14	15	16	17	18	19
No designs eliminated . .	2	2	6	6	14	15	19	25	16	24	17	6	5	6
Elimination Method (*a*) .	3	3	8	9	14	15	16	31	15	22	21	7	7	8
Elimination Method (*b*) .	3	3	8	8	14	13	16	22	13	25	24	8	8	9
Elimination Method (*c*) .	3	3	7	7	13	13	17	22	15	22	20	9	9	8
Method Preferred . . .	any	any	*c*	*c*	*c*	*b* or *c*	*a* or *b*	*b* or *c*	*b*	[illegible] or *c*	*c*	[illegible]	*a*	[illegible] or *c*

The directions for the block-design test, the methods of scoring, and the norms, to which Chapter II of Part B has been exclusively devoted, are based upon the results obtained after rescoring the test blanks, designs XI, XIII, and XVII being eliminated. The designs, after this final revision, have been renumbered as already explained in the introduction to this chapter, each design now being designated by an Arabic instead of a Roman numeral.

(*b*) *Sex Differences*

On the basis of the changes in scoring above indicated, a second analysis was made to determine sex differences.

This attempt was not as thorough or as complete as the first, indications very early pointing toward no change in the original findings. There was no reason to expect any shifts in the comparative performances of the sexes, since all efforts with the designs were in the direction of refining the test procedure and of improving the method for placing a specific intelligence-valuation upon any given response.

The block-design test as it now stands is comparatively free from the disturbing factor of sex for the first ten designs and for ages below thirteen or fourteen. For the last seven designs and for ages above thirteen, boys seem slightly superior to girls. The reason for this difference at the upper ages and for the more difficult designs is not immediately apparent. Attention has already been directed toward reasonable explanations. (See pp. 118–121, and pp. 125–126.)

Possibly introspective evidence might throw considerable light upon the matter. At any rate, the differences between the sexes are not so great as to invalidate the results which one might obtain, nor is a compensating formula necessary to correct any error which might be theoretically assumed. For all practical purposes sex differences may be disregarded.

(*c*) *Final Norms*

The final norms, with a suggested procedure for giving and scoring the block-design tests are presented completely in Chapter II (Part B). The writer has assembled all of this material in one chapter, in order to facilitate the use of the designs and to assist the interpretation of

results derived therefrom. The longer one uses the block design tests, the more is one convinced of the superiority of the P. E. method of scoring and interpretation over the year-scale method. By means of the former, finer gradations are possible, resulting in finer intelligence differentiations. It is therefore recommended that the P. E. method be used in preference to the year-scale method.

CHAPTER V

The Results

1. DO THE BLOCK DESIGNS MEASURE INTELLIGENCE?

To answer this question, one must necessarily first define intelligence. Perhaps the easiest way out of the difficulty is to state that this test measures very largely the same kind of mental processes that are measured by the Binet scale. And if the Binet scale measures intelligence, so do the block designs, as is evident from the high correlation between the two.

To some, this reply may appear an evasion of the issue and quite unsatisfactory. In that case we must review the current definitions of intelligence. These fall practically into two types, (*a*) those expressed by Ebbinghaus, Ziehen, and Meumann, and (*b*) those expressed by Binet, Stern, and Terman. According to the first group, "*Intelligenz ist Kombinationsgabe*" — intelligence is the ability to "put two and two together," the ability to combine, to synthesize. Whenever, for purposes of adaptation or problem solution, we are called upon to bind together, or to relate, two or more experiences which have, in our own lives, never before been tied up, we are utilizing "intelligence" factors in the process. The more efficient this combination in actual practice, or the less open it is to error and misdirection, the more effective has been the functioning of this "intelligence

machinery." There is no question but that "intelligence" is dependent upon a large mass of materials which have been acquired through experience and stored away ready to be called upon whenever occasion demands its use, but "intelligence" as such is neither "experience" nor "information" but "combinative power" — the ability typified, for example, in the mechanics of syllogistic reasoning.

According to the second group, intelligence is the capacity of an individual to *adapt* himself to new situations. Whenever we are confronted with problems, questions, situations of one sort or another, which call for new modes of behavior, an adequate solution of the difficulty, if not accidental, is the result of the operation of "intelligence." Whether "intelligence" has been utilized or not — and if used, to what extent — may be determined by appraising the *result.* What "intelligence" *does* is the matter of chief importance. Only by a knowledge of the range of possible accomplishment and by our ability to measure differences in this capacity can we make practical application of the functioning of intelligence in such fields as education, delinquency, defectiveness, and industry.

It is apparent upon careful analysis that neither of these definitions really touches rock-bottom. Severe in the standards by which definitions of those we examine are judged, we ourselves indulge in an infantile reaction for which we merit some condemnation. As mature, thinking people our definitions should certainly be "better than in terms of use." To define intelligence as "adaptation to new situations" is but little superior to

that of an immature youngster who defined "soldier," "to fight." We wish to know what intelligence *is*, not what it *does*. Although the explanation of intelligence as "combinative ability" falls short of a complete definition, it does introduce us, however, more intimately to the basic processes underlying "intelligence-activity." Psychologically considered, "adaptation" is a most complex activity, and intelligence cannot be defined as adaptation, because we are not here dealing with those elements which condition adaptation.

The reader's attention, at this point, may well be turned to our earlier consideration of the processes of analysis and synthesis. We noted there that these mental activities were of the greatest importance in the growth and development of hierarchies of mental functioning. No matter where our cross-section was made, whether at perception, ideation, judgment, reasoning — there we found these phenomena, analysis and synthesis, explaining to a large extent the compounding and the unifying of processes and of mental material. We noted then that our ability to *analyze* a presented situation, is as important as our ability to synthesize details into a unity. Ebbinghaus, Ziehen, and Meumann overemphasize the importance of "*Kombinationsfähigkeit*," making slight if any mention of "analytic ability." To that extent they fail short of a satisfactory definition of intelligence. To illustrate, suppose one is suddenly, and for the first time, confronted with this problem: Devise an instrument which will assist you in overcoming the effects of gravity. Immediately, numerous ideas come to one's mind. Perhaps you think

of balloons, elevators, aeroplanes, feathers, birds, your body weight, electronic forces, etc., etc. The goal to be attained continually thrusts itself into your consciousness, often working havoc with what seems a perfect plan. Conditions of all sorts are considered, weighed, and measured. Handicaps of all varieties are met with and either they master you, or you master them. Before you have proceeded very far upon your program of action, you have been dealing with a host of details — you have been comparing, discriminating, "analyzing," "combining," judging. Suppose now, some definite solution finally comes to you. Let us assume you have decided upon some intricate electrical mechanism. This decision would never have been reached had various details not been synthesized into what appeared, after a good deal of checking of experiments, a logical, practical plan.

Intelligent acts of all sorts require both an analysis of the situation which confronts one, a *critical* inquiry into methods of solving the problem, and a final synthesizing of details into a consistent whole. Our definition of intelligence as "adaptation" fails to take into account the higher syntheses and analyses of the thinker, the scientist, the generalizer, the discoverer of laws. What factors of "adaptation" are there present in the propounding of a theory of heredity, of a doctrine of evolution, of a "periodic law," of an electronic theory, of a cosmic hypothesis? Nevertheless, such syntheses are often the results of years of painstaking labor, critical search, and the keenest of insight. Few, if any, will deny that these are most decidedly the results of

the highest type of intellectual functioning. It seems clear that the definition of intelligence as "adaptation" breaks down most emphatically at just this point.

Intelligence more adequately defined would be expressed as the ability of an individual to analyze and synthesize. Ability in this direction varies as does physical height. And there can be no intelligent adaption of any sort which is not conditioned upon these fundamental activities. Analysis and synthesis are the mechanics of intellectual behavior — adaptation to changing situations is but *one* of the many results.

To return now to our original proposition, do the block designs measure intelligence? It may be stated in reply that there seems to be no reasonable doubt, substantiation having been obtained through introspections, that these tests require first, the breaking up of each design presented into logical units, and second, a reasoned manipulation of the blocks to reconstruct the original design from these separate parts. The results of this activity, it is presumed, yield a fair index of this analytic-synthetic power which we have termed "intelligence."

2. VALIDITY

To measure the validity of any newly devised test of intelligence is not a simple matter. It devolves upon the standardizer to present evidence that the new intelligence scale measures this inadequately defined entity "intelligence" with approximately the same degree of accuracy as those standards or "measuring rods" now commonly accepted and in current use.

In the following sections the writer will limit his discussion to ten criteria, although admittedly, these are neither exhaustive nor of equal importance:

(1) The mental processes employed;
(2) Increased in score from year to year;
(3) Correspondence of median mental ages;
(4) Correlations between mental ages and intelligence quotients;
(5) Correlations with teachers' estimates of intelligence, and with school standing;
(6) Conformance of intelligence-quotient distribution with normal probability;
(7) Correlations with vocabulary; Trabue B and C and Military Test;
(8) The probable error of a block-design mental age;
(9) The segregation of the feeble-minded;
(10) Percentage ratios of agreement and disagreement with established standards of intelligence.

(1) *Mental Processes Employed*

In devising and standardizing this test the writer did not approach the problem with any bias of "faculty psychology." The idea still seems prevalent, though not as much now as in the immediate past, that in order to possess an adequate measuring instrument for intelligence, the device must contain separate tests for each mental "function": sensation, perception, association, imagination, memory, judgment, reasoning, etc. On the other hand, it has been amply demonstrated that the only intelligence scales worth the name draw service freely from all "functions." Binet has pointed out that

all "intelligent operations" involve the functioning of three primary activities: first, *attention* to the problem presented; second, a conscious attempt on the part of the subject to consummate an adequate *adaptation* to the situation; and third, the exercise of *auto-criticism* in order to determine how efficiently the specific "adaptation" has solved the problem. Put in somewhat different words,[1] "Binet's conception of intelligence emphasizes three characteristics of the thought process: (1) its tendency to take and maintain a definite direction, (2) the capacity to make adaptations for the purpose of attaining a desired end, and (3) the power of auto-criticism." (p. 147.)

A cursory examination of the demands made upon the mental operations of the person tested with the block designs will clearly reveal that *attention*, *adaptation*, and *auto-criticism* are all involved in the successful accomplishment of each task. That point in the graded series of designs at which a child will begin failing to achieve further success, will be a rough measure of the development of his ability to attend, to adapt, and to critically survey his general plan of performance and his ultimate accomplishment. In his discussion of the 'patience test' in the 1908 scale, and these words might as well apply to the block-design tests, Binet states:[2]

"It is a game, but at the same time a work of the intelligence. When one analyzes the operation it is found to

[1] L. M. Terman: "The Stanford Revision and Extension of the Binet-Simon Scale for Measuring Intelligence" (Warwick & York, 1917; 179 p.).

[2] "The Development of Intelligence in Children," Publication No. II (Vineland, 1916).

be composed of the following elements: (1) Consciousness of the end to be attained, that is to say, a figure to be produced; this end must be understood, and kept in mind; (2) the trying of various combinations under the influence of this directing idea, which often unconsciously determines the kind of attempt which should be made; (3) judging the combination formed, comparing it with the model, and deciding if it resembles the other." (p. 195.)

If "intelligence" involves the following mental operations: analyzing, combining, comparing, deliberating, completing, discriminating, judging, criticising, and deciding, then the block-design tests may, with justice, be said to call upon the functioning of intelligence and to that extent they are a measure of that mental capacity.

(2) *Increase in Score from Year to Year*

As regards the second criterion, reference to Graph 6 and to the various tables of Chapter IV will clearly reveal that this requisite is satisfied.

By itself, this evidence would be wholly inconclusive, for the reason that growth in height and in weight also show rising increments with every year of increasing life age, yet these values are not diagnostic of *mental* growth. It is only in relation to other standard criteria that the evidence of yearly increments has any validity.

A slight diversion from the main theme of this section may not be out of place.

A matter worthy of note is the lack of consistence in the variability of one mental age (Binet) in terms of the other mental age (block-design) at various age levels (see Table LXX). Thus up to Binet-age 7, and after

Binet-age 16, the variability in comparable block-design age is slight, but from Binet-age 7 variability increases, reaching a maximum at Binet-ages 10 and 11, and decreasing thereafter to Binet-age 17.

TABLE LXX

BINET-AGE AND BLOCK-DESIGN AGE

Binet-Age

BLOCK-DESIGN AGE	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	TOTAL
Below 5 . . .	3	7	14	14	3	1		1										43
6 . . .				10	10	10	2	3	3	1	1							40
7 . . .				2	7	9	7	2	1			1						29
8 . . .				4	3	4	10	4	2		2	1						30
9 . . .				1	3	14	6	7	1	4	2							38
10 . . .					2	6	5	8	5	2	3		2					33
11 . . .						6	2	4	9	4		1		1				27
12 . . .					1	1	2	6	8	3	4	2	1					28
13 . . .							1	3	7	5	3	1	1	1				22
14 . . .							2	3	2	7		4		2				20
15 . . .								2	3	5	3	4	1					18
16 . . .								1	2		1	1	1	2		1	3	12
17 . . .									1			2	4	1	3	2	1	14
18 . . .											1				2			3
19 . . .													1	2			1	4
20 . . .														2	2	1		5
Total .	3	7	14	31	29	51	37	44	44	31	20	17	11	11	7	4	5	366

(Note: Age 10 means 10–7 to 11–6, etc.)

Rather than regarding this condition a handicap, invalidating the tests as a differentiating mechanism for diagnosing feeble-mindedness, it appears, on the contrary, to possess distinct advantages. Too frequently do the Binet tests leave us uncertain as to definitemental classification after obtaining a quotient of 73 or 76. Other evidence in the case, educational, social, biological,

industrial, medical, is either meager or inconclusive. We wonder whether the child's superiority in memory, for example, may not account for a higher quotient than he probably actually possesses. At other times, we are at a loss to account for exceptional success in one or more of the tests in higher ages, unless it be explained on the basis of previous familiarity. It is at such times that an additional differentiating mechanism, such as the block-design tests, may be of pronounced value in forming a more definite decision as to further social and educational handling.

Suppose, for example, we have just concluded the examination of a child whose life age is 15 years, and who obtains a Binet mental age of 11. The I. Q. in this instance would be 74. If then the block-design tests are given, on the basis of Table LXX, this Binet 11-year-old may theoretically obtain a mental rating anywhere between 6 years and 17 years on the block-design tests — a possible divergence of five years in either direction.

The value of this apparent disagreement will be self-evident to psychoclinicians who are daily confronted with the need for rapid, yet accurate, prognostic diagnoses. In the above-mentioned instance, an I. Q. of 50 on the block-design tests coupled with a Binet I. Q. of 74 and a rather negative history of self-dependence might be of unquestioned assistance in outlining a rational plan for further care and direction of a subject under examination. The same value, of course, would attach to a block-design test-result in the case of a child with a Binet I. Q. of 74 and with a block-design I. Q. of 150.

In view of a later return to this same matter, the diagnostic importance of this divergence will not be stressed any further at this point.

(3) *Correspondence of Median Mental Ages*

At each life age do the median mental ages obtained by the block-design tests correspond with the median mental ages obtained by the Binet tests? This question is an important one, and the extent of correspondence or deviation should measure very largely the validity of the newly devised tests.

In the following table this comparison is presented:

TABLE LXXI

Correspondence of Median Mental Ages

Life Age Yrs.	No. of Cases	Median Binet Age Yrs.-Mos.	Median Block-Design Age Yrs.-Mos.	Difference Between Medians (Mos.)	Average of Two Medians Yrs.-Mos.
6 . .	16	6– 1	5– 3	10	5– 8
7 . .	27	7– 5	6– 8	9	7–½
8 . .	28	8– 0	8– 2	2	8–1
9 . .	27	9– 0	8–11	1	8–11½
10 . .	33	9– 8	10– 3	7	9–11½
11 . .	30	10– 6	11– 6	12	11– 0
12 . .	29	11–10	12– 5	7	12– 1½
13 . .	32	11– 5	12– 6	13	11–11½
14 . .	28	13– 9	13– 6	3	13– 7½
15 . .	19	13– 3	14– 0	9	13– 7½
16 . .	16	15– 6	13–10	20	14– 8
				Average—8.5	

Four important items are worthy of note: In the first place, the average deviation of the median Binet ages from the life ages at each year is 6 months; second, the average deviation of the median block-design ages from

the life ages at each year is 8.8 months; third, the average deviation between the two intelligence-test medians is 8½ months, and finally, *the arithmetic mean of the two medians for each life age results in a more accurate approximation of what may be the "true" mental age than either median taken alone.*

We have, perhaps, in our early enthusiasm over the diagnostic accuracy of the Binet scale overlooked the importance and the need for calling to our assistance supplementary information which should aid us in arriving more accurately at a positive diagnosis. The further our studies upon the fallibility of intelligence ratios continue [1] the more does it become apparent that, at least for the determination of mental deficiency, a Binet I. Q. alone is a hazardous symptom upon which to base a definite diagnosis of feeble-mindedness. Many and frequent have been the criticisms, sometimes mild, at other times venomous, that the accuracy of the Binet result was impaired by the intercipient factors of schooling and linguistic facility. It may, consequently, not be a mistaken procedure to utilize both a language and a performance test in the determination of mental level — the performance test to compensate for the overemphasis in the Binet scale of certain capacities not highly correlated with intelligence *per se*, the Binet scale to compensate for the inevitable easy-spread of ability characteristic of practically all performance intelligence-tests.

[1] Florence Mateer: "The Diagnostic Fallibility of Intelligence Ratios," *Ped. Sem.*, 1918, 25: 369–392.

E. A. Doll: "The Growth of Intelligence," *Psychol. Monographs*, 1921, Vol. 29. No. 2, 130 p.

After comparing the Binet age medians with the block-design age medians for each life age (Table LXXI), it may be sufficient to remark, in conclusion, that the approximation between the Binet and the block-design medians is remarkably close, especially when we consider that the block-design tests are free of the "language factor" and of the influence of schooling.

(4) *Correlation between Mental Ages and Intelligence Quotients*

One of the greatest handicaps experienced in standardization is to select a standard of comparison which is absolute. All existing standards are, unquestionably, affected by a variety of disturbing factors, frequently difficult to determine, and always difficult to discount. As in the year-scale standardization, the writer may have erred in selecting the results of Binet testing as a criterion of accurate mental age. That procedure, however, was necessary because we have, as yet, no other equally accurate measuring stick — one which has attained a similar degree of general confidence.

For the same reason, the Binet results are again utilized to check the validity, in the present instance, of the block-design ages.

It may be of interest, first, to state the separate correlations of the two mental ages with *life* age. Conformance with life age is, apparently, greater with the Binet than with the block-design tests.

1. The correlation between Binet age and life age is $+.80$ (P. E. $\pm .01$) (291 public school cases).

2. The correlation between block-design age and life age is + .66 (P. E. ± .02) (291 public school cases).

Segregating the results for the total 366 cases, for the 291 public school cases, and for the 75 feeble-minded the correlation between Binet age and block-design age follows:

3. The correlation between Binet age and block-design age is + .82 (P. E. ± .01) (366 cases).
4. The correlation between Binet age and block-design age is + .81 (P. E. ± .01) 291 public school cases).
5. The correlation between Binet age and block-design age is + .67 (P. E. ± .05) (75 feeble-minded cases).

The lower correlation between Binet and block-design ages for the feeble-minded is sufficiently great to require some comment.

The application of the tests to the feeble-minded was purposely divided, in point of time: the first group of feeble-minded cases (28) were examined at Vineland, the summer of 1917; the second group (47) at the Sonoma State Home (Eldridge, Cal.), were tested during the spring of 1918. Almost one year later, because of accumulated experience with the block-design tests the outstanding deviations in performance of some of the feeble-minded could be more intelligently interpreted. Dr. George Ordahl, the psychologist of the Sonoma State Home, remarked during the period of experimenta-

tion that the block-design tests seemed to measure *functional efficiency* to a far greater degree than did the Binet tests. Looking back to the writer's personal knowledge of the cases at Vineland this remark was corroborated. It appears that many feeble-minded actually work at a higher level of complexity and efficiency than one would expect on the basis of Binet mental age. For the reason that the Binet possibly underestimates the work-capacity of certain timid or retiring or inexpressive types of defectives, or for the reason that the block-design tests possibly place less emphasis upon verbal expression and more upon performance, it seems that in certain instances the feeble-minded obtain inconsistent scores on the two scales, yet their ratings by the block-design tests are a fairer index of their functional efficiency in various occupations than the Binet.

This is a matter which can well afford further investigation. There still remain unchartered territories without number awaiting exploration by those interested in determining individual differences among the feeble-minded as well as the differences between the normal and the feeble-minded.

As regards the correlation between the I. Q.'s of the Binet and the block-design tests, the results again are segregated for the total 366 cases examined, for the 291 school children, and for the 75 feeble-minded. These follow:

6. The correlation between Binet I. Q. and block-design I. Q. is + .80 (P. E. ± .01) (366 cases). The table is herewith presented:

TABLE LXXII

BINET I. Q. AND BLOCK-DESIGN I. Q.

Binet I. Q.

Block-Design I. Q.	20	30	40	50	60	70	80	90	100	110	120	130	140	TOTAL
30 . .		13	7	1	1	1								23
40 . .	1	3	4	6	2	3		1						20
50 . .		1	1	5	7	1	1	2						18
60 . .			2	2	4	8	10	1	2					29
70 . .				1	1	4	6	4	3	2				21
80 . .					1	3	11	12	10	5	2			44
90 . .				1	1	3	9	10	20	2	2			48
100 . .						1	6	20	20	7	1	1		56
110 . .						2	7	10	13	7	2	1		42
120 . .							1	5	8	8	5	1	1	29
130 . .								1	3	4	5	2		15
140 . .								2	3	3	5	1	1	15
150 . .									1	1				2
160 . .										1	1		1	3
170 . .											1			1
Total .	1	17	14	16	17	26	51	68	83	40	24	6	3	366

(Note: I. Q. of 50 means 46–55, etc.)

7. The correlation between Binet I. Q. and block-design I. Q. is + .58 (P. E. ± .03) (291 school children).
8. The correlation between Binet I. Q. and block-design I. Q. is + .67 (P. E. ± .05) (75 feeble-minded cases).

The correlations of the two I. Q.'s, both for the total 366 cases and for the 75 feeble-minded are equal to the comparable correlations of the two mental ages mentioned above. In the case of the 291 school children there is a marked difference, however. Whereas the correlation between Binet age and block-design age is + .81, the correlation between Binet I. Q. and block-

design I. Q. is only + .58. It is difficult to explain this discrepancy, both tables having been checked for accuracy on three different occasions. The only plausible explanation at this time is the greater range of the mental age table as compared with the I. Q. table. The latter has fewer class-intervals, I. Q.'s having been taken in ranges of ten each.

In this connection may also be presented the individual deviations in I. Q. between the Binet and the block-design tests. No one would expect exact correspondence, yet a high percentage of agreement should be demanded. In the following tables, segregated for the 291 public school children and for the 75 feeble-minded, this correspondence is presented.

TABLE LXXIII

CORRESPONDENCE OF I. Q.'S

(291 P. S. Cases)

I. Q.	Frequency		Total
	B. D. I. Q. Higher	B. D. I. Q. Lower	
Same			7
1–5	27	29	56
6–10	35	22	57
11–15	20	28	48
16–20	25	22	47
21–25	16	11	27
26–30	12	5	17
31–35	6	6	12
36–40	3	[illegible]	5
41–45	5	3	8
46–50	3	1	4
51–55	[illegible]		[illegible]
56–60	1		1
Median I. Q. Difference .	15	13	14

From the above table it is evident that 50 per cent of the cases show no greater deviation than 14 I. Q. between the two tests. It is also clear that there are almost as many lower deviations as higher:

The total number higher	— 155 cases	(53%)
The total number lower	— 129 cases	(44%)
The total number the same	— 7 cases	(3%)

If anything, the block-design tests err in the direction of leniency rather than severity of mental age estimation.

The closeness of correspondence is again apparent from the following analysis:

Total deviations no greater than	5 I. Q.	63 cases	(22%)
" " " " "	10 I. Q.	120 cases	(41%)
" " " " "	15 I. Q.	168 cases	(58%)
" " " " "	20 I. Q.	215 cases	(74%)
" " " " "	25 I. Q.	242 cases	(83%)
" " " " "	30 I. Q.	259 cases	(89%)
" " " " "	35 I. Q.	271 cases	(93%)

For the feeble-minded, the data runs parallel, [1] except that to a certain extent the figures are much more significant in comparison with the same group of data for the 291 P. S. cases, and from the point of view of the diagnostic efficiency of the block-design tests. So in Table LXXIV the correspondence of I. Q.'s for the 75 feeble-minded cases is presented:

[1] It should be stated that all block-design zero scores were given a mental age rating of 5 years 3 months. This inaccuracy increases the I. Q. deviations. But this was purposely done in order artificially to increase the differences between the two I. Q.'s wherever possible. The comparisons throughout are, therefore, quite conservative.

TABLE LXXIV

CORRESPONDENCE OF I. Q.'s

(75 F. M. Cases)

I. Q.	Frequency		
	B. D. I. Q. Higher	B. D. I. Q. Lower	Total
Same			3
1– 5	15	13	28
6–10	5	13	18
11–15	6	3	9
16–20	[illegible]	3	5
21–25	1	2	3
26–30	1	3	4
31–35	2	[illegible]	4
36–40		1	1
Median I. Q. Difference .	6	8	7

It is clear from this table that 50 per cent of the total feeble-minded examined show no greater deviation than 7 I. Q. between the two tests. This similarity of results indeed merits special emphasis. It will also be noted that whereas, for the average school child, there is a strong tendency to secure a somewhat *higher* score on the block-design tests, this is reversed for the feeble-minded. The latter group tends to secure lower scores. This is evident from Table LXXIV, and also from the following totals:

The total number higher — 32 cases (43%)
The total number lower — 40 cases (53%)
The total number the same — 3 cases (4%)

This evidence is in harmony with the remark previously made, that the block-design tests assist in the

differentiation of the feeble-minded from the border-zone or the normal case.

Although the above-mentioned differences exist between the two I. Q.'s, yet correspondence is higher for the feeble-minded than for the public school group:

Total deviations no greater than	5 I. Q.	31 cases (41%)
" " " " "	10 I. Q.	49 cases (65%)
" " " " "	15 I. Q.	58 cases (77%)
" " " " "	20 I. Q.	63 cases (84%)
" " " " "	25 I. Q.	66 cases (88%)
" " " " "	30 I. Q.	70 cases (93%)

In conclusion, it may be stated that the correlations between the two test methods, considering their marked difference in character, are high enough to guarantee frequent correspondence between their mental ages and I. Q.'s.

(5) *Correlations with Teachers' Estimates and with School Standing*

The search for the "fountain of youth" presented few if any greater difficulties than the search for criteria by means of which to weigh the validity of measuring instruments for intelligence. It is comparatively easy to devise tests, to develop an efficient technique for application, to establish norms, and to obtain results which may possess some psychological, educational or social significance. But to find a stable criterion for establishing validity appears almost hopeless.

It has been customary to rely upon teachers' estimates of intelligence to check the validity of one's intelligence-test schema. It is valuable to utilize this "common-sense" check, even though high correlations between teachers' judgments and test results cannot be expected,

and in spite of the fact that there are some very serious forces invalidating such judgments. It is clear that the same high I. Q. receives a lower estimate by the teacher if a child is in upper grades rather than in the lower grades, because of the elimination of poorer pupils in the upper elementary grades. Thus an I. Q. of 90 is more likely to receive an estimate of average or above average in the lower grades, whereas in the upper grades or in the high school the same I. Q. will more probably receive an estimate of below average or inferior. Other writers have already indicated other factors, having no bearing upon or relation to intelligence, which interfere with the accuracy of judgment on the part of teachers.

School marks, also valuable criteria with which to check the results of a newly devised test, are also inefficient to an extent, because they are generally crude, are limited to a rather narrow range, contain spurious elements such as neatness, order, and so on, are often given for disciplinary reasons, or as bait for an educational spurt, and are not based upon uniform standards.

However, there is a corroborative value in teachers' estimates and in school standing which is worth mentioning.

9. The correlation between teachers' estimates of intelligence and Binet I. Q. is + .47 (P. E. ± .03) (291 school children).
10. The correlation between teacher's estimates of intelligence and block-design I. Q. is + .23 (P. E. ± .04) (291 school children). The table is herewith presented.

TABLE LXXV

TEACHERS' ESTIMATES AND BLOCK-DESIGN I. Q.

Teachers' Estimates

Block-Design I. Q.[1]	4	5	6	7	8	9	10	TOTAL
30			1					1
40				1				1
50			2	1	1			4
60	2	2	6	5	3	1		19
70		1	5	9	3			18
80		1	6	21	10	3	1	42
90		1	13	20	9	1		44
100		3	7	26	14	5		55
110	2	1	7	21	8	3		42
120		2	2	13	9	3		29
130			2	5	4	3	1	15
140				5	6	3	1	15
150				2				2
160			1	1	1			3
170				1				1
Total	4	11	52	131	68	22	3	291

[1] (30 signifies scores between 26 to 35, etc.)

$$r = + .23, \pm .04$$

Although the correlation between block-design age and Binet age is + .82, teachers' estimates of intelligence correlate only one half as much with the block-design I. Q.'s as with the Binet I. Q.'s. The reader may recall that one of the original objections to the Binet scale was that it measured school training. Only to a limited extent has this been denied, the explanation having been made that the tests measure intelligence through the medium of knowledge only partly influenced by school training. It has been admitted, true, that practically all children are exposed to these educational influences, but the ultimate difference in achievement is explainable on

the basis of differences in endowment. However this may be, the results of the block-design test would perhaps tend to show that there is more to this charge than we have been inclined to admit. It will, no doubt, be conceded without much question that the block-design tests are less affected by school training than the Binet.

Regarding the correlations between school standing (school marks) and block-design I. Q. the following are the results:

11. Block-design I. Q. and Reading: + .18, P. E. ± .04 (276 cases).
12. Block-design I. Q. and Language: + .20, P. E ± .04 (259 cases).
13. Block-design I. Q. and Household Arts: + .22 P. E. ± .09 (48 cases: 44 girls, 4 boys).
14. Block-design I. Q. and Drawing + .25 P. E. ± .04 (258 cases).
15. Block-design I. Q. and Arithmetic: + .31, P. E. ± .04 (251 cases).
16. Block-design I. Q. and Manual Training: + .36, P. E. ± .08 (58 cases: 56 boys, 2 girls).

It will be observed that block-design results correlate least with reading and language, and most with manual training and arithmetic. When discussing sex differences in an earlier section of this monograph, attention was called to the slight superiority of boys over girls at the higher ages. It was mentioned then, that perhaps the mechanical nature of some of the higher designs gave boys a slight advantage. The evidence here is, to an extent, a corroboration of this view. This condition

does not point necessarily to a weakness in the test scheme. It will be conceded by those who analyze the activity that success in manual training involves the utilization of intelligence in planning one's work, in criticising one's progress through the various stages of accomplishment, and in holding the goal to be achieved in mind and in attaining it. It would possibly be generally admitted that efficiency in manual training draws no *less* upon "intelligence" than do reading and language.

It may also be of interest to compare the average of the correlations between reading, language and arithmetic, the "mind" subjects on the one hand, and household arts, drawing, and manual training, the "hand" subjects on the other. The averages of the two groups are + .23 for the first subjects, and + .28 for the second. Again we find corroboration of the contention that the block-design tests are actually tests of *performance*, more free of the language factor than many of the tests, necessarily dependent upon language, now in common use. The need for the application of such a test as this in combination with others, in order to estimate accurately a child's intelligence, appears self-evident.

When the teachers were requested to fill out the questionnaire they were asked to answer this question for each child examined: "If this child were given a mental test involving chiefly the *use of language*, or the ability to understand and handle language symbols (as contrasted with a test requiring silent observation and *performance*) would you expect him (her) to pass; very low, low, below average, average, above average, high,

very high?" It was expected that the replies to this question would bring to light those cases where undue superiority or undue inferiority in language understanding and expression might affect the results of a mental test dependent upon language symbols for comprehension and success. The correlation between block-design I. Q. and these estimates by the teachers follows:

17. Block-design I. Q. and teachers' estimates of language vs. performance: $+ .24$, P. E. $\pm .04$ (291 cases).

Analyzing the table of data it is found that 6 per cent of cases possessing I. Q.'s of 106 and over were estimated by the teachers likely to obtain very low, low, or below average scores on a "language" intelligence test, and at the other extreme 10 per cent of cases possessing I. Q.'s of 95 or under were estimated by the teachers likely to obtain above average, high, or very high scores on a "language" intelligence test. In other words, according to the judgment of teachers, it would appear that a total of 16 per cent, or one sixth of all children in the average public school, are likely to be misrated if judged by a "language" test alone. This proportion is probably too high. At any rate, the data is offered for what it is worth.

By way of summary, it may be remarked that not only do the block-design tests show a reasonable correlation with school standing and teachers' estimates of intelligence, but the emphasis upon performance as opposed to verbal impression or expression is a desirable feature of the test.

(6) *Conformance of Intelligence Quotient Distribution with Normal Probability*

A very necessary index in weighing the validity of any standardized test is to determine the extent to which an actually found distribution conforms to its theoretical distribution.

The assumption that the distribution of intelligence follows normal probability is recognized as an arbitrary one, but it is utilized merely as a working hypothesis. If not this, what else? Until further investigation gives us more light on this problem we must continue to grope, more or less, in the dark.

In the following table are presented the I. Q.-range distributions for the Binet and the block-design tests. The respective percentage values are compared with what one should theoretically expect.

Attention is again called to the "spread" of ability greater for the block-design tests than for the Binet, previously mentioned. It is apparent that "intelligence" measured by the Binet test manifests a lesser diveristy than the block-design test. Which of the two is more truly representative it is difficult to state dogmatically. Perhaps an average of the two approaches more nearly the truth. It may be of interest to mention a note contained in a letter to the writer from W. A. McCall in which he remarks "that the distribution of Reading Quotients," obtained from his recently completed reading scale, "conforms more closely to your I. Q.'s than Terman's."

Whatever the true situation regarding the general character of the distribution of intelligence, the fre-

TABLE LXXVI

INTELLIGENCE QUOTIENT RANGES

	26 TO 35	36 TO 45	46 TO 55	56 TO 65	66 TO 75	76 TO 85	86 TO 95	96 TO 105	106 TO 115	116 TO 125	126 TO 135	136 TO 145	146 TO 155	156 TO 165	166 TO 175
Stanford Binet Obtained . . .				1.7	5.5	16.5	22.7	28.2	13.8	8.3	2.1	1.0			
Theoretical Expectation				.16	1.6	8.5	23.42	32.64	23.42	8.5	1.6	.16			
Block-deisgn Obtained . . .	.034	.034	1.4	6.5	6.2	14.4	15.1	18.9	14.4	10.0	5.2	5.2	.07	1.0	.034
Theoretical Expectation (Median at 99) . . .	.49	1.28	2.78	5.21	8.67	12.05	14.66	15.30	13.84	10.69	7.25	4.11	2.08	.88	.33

quencies of the block-design I. Q.'s conform more closely to theoretical expectation. Thus the average deviation from theoretical expectation for the Binet I. Q. ranges is 3.3 per cent per I. Q. group. The average deviation for the block-design tests is only 1.4 per cent per I. Q. group.

In conclusion, the evidence of correspondence between the empirical and the theoretical distributions testifies to a fair degree of validity of block-design I. Q.

(7) *Correlations with Vocabulary, Trabue B and C, and Military Test*

Mention has already been made of the correlations between block-design and Binet test results. The data of the Stanford-Binet vocabulary, the Trabue Language Scales B and C, and the military test were also utilized to verify the accuracy with which the block-designs estimate intelligence. These criteria were especially chosen for the reason that, besides the inevitable differences to be expected from the application of two intelligence scales, additional disturbing factors of language and group testing were added in order to exaggerate the inherent discrepancies. It was also the intention to obtain some information which would throw light upon the much mooted questions, "Does the Binet test exaggerate the importance of language in contradistinction to performance?" and "Are there specialized capacities of 'doing' which are as indicative of intelligence as those concerned with 'talking' and 'information'?" This problem, however, can only be barely touched upon at this point.

In the following tables are presented, first, the correlations between Stanford-Binet mental age and Stanford-Binet vocabulary score, and second, between block-design score points and Stanford-Binet vocabulary score:

TABLE LXXVII

CORRELATION BETWEEN BINET AND VOCABULARY

VOCABULARY SCORE*	BINET MENTAL AGE														TOTAL
	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
0–4	4	4													8
5–9	3	7	4	1		1									16
10–14	6	4	9	1	1										21
15–19		10	19	6	1										36
20–24			11	16	13	3									43
25–29			3	11	13	2	2								31
30–34			2		11	12	5	3							33
35–39				1	3	11	4	1	1	1					22
40–44					1	10	7	4	1						23
45–49					1	1	6	5	4	1					18
50–54						2	6	4	7	4	2	1			26
55–59								2	2	2	2				8
60–64						1		1	1	2	2	2			9
65–69									1		2	3	2		8
70–74											2	1	1	1	5
75–79													1	3	4
80–84															
85–89										1					1
90–94											1			1	2
Total	13	25	48	36	44	43	30	20	17	11	11	7	4	5	314

$r = +.91$, P. E. $\pm .007$

* These class-intervals, although not the best, were used in order to correspond with Terman's.

It may be worth remarking at this point that of the three children who achieved the highest scores on the vocabulary test, two were inmates of the institutions for the feeble-minded! On the block-design test, however, their mental incapacity is evident, one achieving a score

between 21–30, the other a score between 71 to 80. The Binet ages of the two fell, for the first, between 14–7 and 15–6, and for the second, between 15–7 and 16–6.

Referring to Table LXXVII, the correlation ratio + .91 is the same as that found by Terman in his analysis

TABLE LXXVIII

CORRELATION BETWEEN BLOCK-DESIGN AND VOCABULARY

Vocabulary Score	Block-design Score Points															Total
	0	1 to 10	11 to 20	21 to 30	31 to 40	41 to 50	51 to 60	61 to 70	71 to 80	81 to 90	91 to 100	101 to 110	111 to 120	121 to 130	131 over	
0– 4 . .	35	18	4	3												60
5– 9 . .	2	9	2	1	1	1										16
10–14 . .	5	10		5	1											21
15–19 . .		14	10	8	3	1										36
20–24 . .	1	14	14	3	4	3	1	1	1			1				43
25–29 . .		5	6	8	2	3	5		2							31
30–34 . .		5	2	4	4	5	2	5	2	3		1				33
35–39 . .		[illegible]	1	2	2	2	2	3	3	1	2	1		1		22
40–44 . .		2	4	4	2	2	1	2	1	1	2	1	1			23
45–49 . .			2	2	1	5	3		2	1		1		1		18
50–54 . .		1		[illegible]	3	3	1	3	2	3	4	2	2			26
55–59 . .			1	1				1	1		2	2				8
60–64 . .						1	1		1	1		2	2		1	9
65–69 . .			1									[illegible]	1	4		8
70–74 . .												2		1	2	5
75–79 . .												2	1	1		4
80–84 . .																
85–89 . .				1												1
90–94 . .									1			1				[illegible]
Total .	43	80	47	43	23	26	16	15	16	10	10	18	7	8	3	365

$r = +.77$, P. E. $\pm .02$

of the records of 631 school children,[1] wherein mental age and vocabulary score were correlated. Based upon statistical analysis, not personal opinion, the vocabulary

[1] L. M. Terman: "The Intelligence of School Children" (1919), p. 310.

test stands out clearly as the best single test in the whole Binet series. The correlation ratio + .91 is to a certain extent spurious because vocabulary score entered into the estimation of mental age and to that extent weights the Binet mental age in its favor. In view of this condition the correlation between Binet mental age and block-design mental age (+ .84) is not so much lower than one would conclude off-hand. Then, again, reference to Table LXXVIII and to Terman's Table 41[1] reveals clearly that the variability of a block-design score in terms of Binet mental age is not so great, when we observe that the vocabulary test presents approximately the same degree of variability, and in spite of the fact that the vocabulary aided in the determination of Binet mental age. For 482 miscellaneous adults — hoboes, prisoners, delinquents and business men — Terman found the correlation between mental age and vocabulary to be + .81, three points lower than the correlation between the Binet and the block-design tests.

It is quite surprising to note the correspondence between efficiency in the block-design tests and in vocabulary, the former apparently concerned with hand-manipulations, the latter with language, supposedly affected by schooling and experience. In the past we have been inclined to look to performance tests as "fillers-in" for the assumed inadequacies of the Binet tests. The high correlation between the two types of examination leads one to the conclusion that both systems call upon somewhat similar mental processes — processes involved in the normal functioning of intelli-

[1] L. M. Terman: "The Intelligence of School Children" (1919), p. 310.

gence. So that, although the two methods of expressing this mental power differ radically in character, the mental manipulations making for successful achievement are fundamentally the same.

An analysis of the reactions of the feeble-minded to the two test systems reveals roughly three rather distinct groups. On the one hand there are what Binet has termed the "loquacious types." These are the feeble-minded who have an easy facility with language, and who, because of their loquaciousness and ability to speak with apparent intelligence on many subjects, give to the lay mind the impression of average mental capacity, a sort of general ability and "intelligence." These aments secure high scores in the vocabulary test, but their performance in the block-design tests is strikingly low. Then again, there are a number of others, considerably backward in speech expression and in language ability — what some have designated "the silent type." The character of their actual daily activity belies their mental rating on the Binet scale. Low in their achievements on this intelligence test, they are nevertheless capable of considerable rather intelligent effort if taught to *do* something, and are placed upon their own initiative which will involve no greater than a few minor adaptations. These mental deficients although markedly unintelligent on a test involving the understanding and the use of language, nevertheless are capable of a considerable amount of useful and complex effort on the institution grounds. Those who have actually worked with the feeble-minded in their institutions will readily recall numerous instances which are corroborative of

the above remarks. Finally, there are those aments who manifest the same degree of intelligence-development by both tests. This group, of the three, is apparently the largest.

It appears that our earlier notion that language capacity and performance were two somewhat mutually exclusive abilities was to a large extent erroneous. Granted a certain level of intelligence-development, the vocabulary test or the Binet scale are, in most cases, no less effective measures of that level than a reliable performance test itself, no matter how carefully or efficiently devised or standardized. But, in case of doubt, both tests may be given to advantage, averaging the results of the two.

What has already been said regarding the relation between the diagnostic efficiency of a language test and a performance test holds true for such a language test as Trabue B and C. Devised originally to measure pedagogical achievement in language, it was found very early to be also diagnostic of intelligence-development. It is again surprising to note the relatively close correspondence in scores between this language and the block-design tests. See table on the next page.

The correlation of: + .57 between these two test schemes is indicative of a certain amount of common mental activity characteristic to both, irrespective of outward differences in the physical form of the tests themselves.

An analysis of the regression of the block design on Trabue score may be of interest. From the table it is evident that as Trabue score increases, the increase

in block-design score is relatively slight. When, however, the middle range of Trabue scores are reached the increase in one is paralleled by similar increases in the

TABLE LXXIX

CORRELATION BETWEEN TRABUE B AND C AND BLOCK-DESIGN SCORES

TRABUE SCORE B AND C	0	1 TO 10	11 TO 20	21 TO 30	31 TO 40	41 TO 50	51 TO 60	61 TO 70	71 TO 80	81 TO 90	91 TO 100	101 TO 110	111 TO 120	121 TO 130	131 and over	TOTAL
0– 1 . .	7	18	5	1	1											32
2– 3 . .		1														1
4– 5 . .	1	1			1											3
6– 7 . .		1	2	1												4
8– 9 . .		3	1	4	1											9
10–11 . .	1	8	2			1										12
12–13 . .		7	1	5	2		1		1							17
14–15 . .		4	2													6
16–17 . .		4	3	3	1	2	1	1	1							16
18–19 . .		1	2	5	2	1	1	2								14
20–21 . .		3	4	8		2	3	1	1	2		1				25
22–23 . .		1	4	3	2	5	1	3	1	2	3	1				26
24–25 . .		2	1	1	2	4	2	3	3	1	1		1			21
26–27 . .			5	4	5	3	4	1	3	3	2	3			1	34
28–29 . .		1		1	2	4		1	2			2	2	3		18
30–31 . .					1	3		2	1		2	3	2	2		16
32–33 . .				1			1			2						4
34–35 . .					1			1				4	1	2	1	10
36–37 . .												1	1	1	1	4
38–39 . .												1				1
Total .	9	55	32	37	21	25	14	15	13	10	8	16	7	8	3	273

$r = + .57$, P. E. $\pm .03$

other, except that the *higher* the Trabue score the greater the variability in terms of block-design score. So *vice versa*, the *lower* the block-design score the greater the variability in terms of Trabue score. This condition may be interpreted as indicating that at the lower ages facility in language is rather low, so that block-design scores are higher at any given age. On the other hand,

the reverse is true for the higher ages: Trabue scores here seem to possess higher diagnostic efficiency.

A number of the children who were included among those tested for block-design ability had been given the military test when it was in its early stages of standardization. This again is a test in which the "language factor" plays no insignificant part, yet is a recognized instrument for intelligence estimation. The cases which could be included in this analysis were limited, but the results may be of interest. The table follows:

TABLE LXXX

CORRELATION BETWEEN MILITARY AND BLOCK-DESIGN SCORES

Military Test Score	Block-design Scores															Total
	0	1 to 10	11 to 20	21 to 30	31 to 40	41 to 50	51 to 60	61 to 70	71 to 80	81 to 90	91 to 100	101 to 110	111 to 120	121 to 130	131 and over	
0– 19	2	9	3	5	1											20
20– 39		2	4	4												10
40– 59		1	5	5	1	1						1				14
60– 79		1		3	1	1			1		1					8
80– 99		2	1	3		2	1									9
100–119		2	1	1		1			1			1				7
120–139			3		3	2		2		1						11
140–159				2	3	6	1	2	4		1					19
160–179			1	2			5	3	2	2	1	1				17
180–199			1	1	2	2	1		2		1					10
200–219			1	1				2	2			3	3	1		13
220–239						1	2			1	1	2		1		8
240–259			1					1				2		2	2	8
260–279												2	2	3		7
280–299												1	2			3
300–319															1	1
320–339												1				1
340–359																
360–379												1				1
Total	2	18	20	27	11	16	10	10	12	4	5	15	7	7	3	167

$r = +.71$, P. E. $\pm .02$

The correlation + .71, between the military test and the block-design test, although surprisingly high to those accustomed to regard "performance" and "linguistic ability" as separate and distinct capacities, is nevertheless to be expected in light of the fact that both test schemes aim to measure intelligence.

In the following tabular arrangement the correlations of the block-design test with the three tests in which "language" plays an important rôle are given:

1.	Block-design test and Binet Vocabulary .	+ .77, P. E. ± .02
2.	Block-design test and Trabue B and C .	+ .57, P. E. ± .03
3.	Block-design test and Military	+ .71, P. E. ± .02

In order to determine whether the correlations between the block-design test and the three tests designated as involving the use of language might have been higher but for the interference of this "language" factor, the scores of the Binet vocabulary and the military test were correlated, both having the "language factor" in common. It is presumed that if these latter two tests show no greater correlation than either test with the block-designs, that other factors than language facility enter to lower the correlation from + 1.00.

The correlation between Binet vocabulary and military test is + .87, P. E. ± .02 (167 cases).

It is, consequently, apparent that other disturbing factors than "language" are operative tending to reduce the correlations between these three tests. The correlations between the block-design and any one of the three other tests is not very appreciably lower than the correlations between any two random "language" intelligence tests.

This brings us to our original problem, "Does the Binet test exaggerate the importance of language in contradistinction to performance?" If such a condition does exist, its importance has been greatly overemphasized in the past. And to our other question, "Are there specialized capacities of 'doing' which are as indicative of intelligence as those concerned with 'talking' and 'information'?" a rational reply based upon our statistical data would include the remark that there are no such compartment-like functions, and that facility in one direction is highly correlated with facility in the other.

From the point of view of the validity of the proposed test-scheme, the block-design tests, the above-mentioned correlations between this performance test and typical language-intelligence tests bear out the contention that the block designs measure intelligence rather effectively.

(8) *The Probable Error of a Block-design Mental Age*

The consistency with which a newly-devised measuring instrument estimates presented values may be regarded a fair index of its validity. Extreme variability and haphazard deviations from normal expectation may be interpreted as indicating lack of efficiency in evaluating quantitative differences. On the other hand, repeated identity of determined values for any given objective quantity argues for proficiency of the measuring tool. It is, therefore, regarded important to determine to what extent, and with what probability, extreme deviations of significant magnitude may be expected.

The probable error (P. E.) is the statistical device whereby such consistency may be computed.

In his study of vocabulary, Terman found[1] the probable error of a mental age based upon vocabulary score alone to be 9.6 months (631 children). In other words, based on vocabulary *alone*, mental age thus determined will not deviate from the mental age derived from applying the entire Stanford-Binet scale more than 9½ months in 50 per cent of cases. The table of ratios follows:

Deviation	*Per Cent of Cases*
9½ mos.	50
12 mos.	40
18 mos.	20
24 mos.	10
36 mos.	1

$$r = + .91$$

A difference in mental age as great as three years may be expected in only one case out of 100, and a difference of two years in only 1 out of 10. Or, put differently, taking any obtained score, the chances that it is within one year correct (one year above or one year below) are 60 that it is, to 40 that it is not. The chances that the score is within two years of the correct mental age are 90 to 10, and within three years, 99 to 1.

For adults the vocabulary test did not function as efficiently. The probable error of a mental age based

[1] Terman, Lewis M., Kohs, S. C., and others: "The Vocabulary Test as a Measure of Intelligence," *Journal of Educ. Psychol.*, October, 1918

upon the vocabulary score with this group of cases was 12 mos. The table of ratios follows:

Deviation	*Per Cent of Cases*
12 mos.	50
18 mos.	31
24 mos.	18
36 mos.	4.3

$r = + .81$

A divergence greater than one year between mental age as determined by the vocabulary alone and that of the Binet scale applied as a whole was to be expected in 50 cases out of 100. A two years' divergence or more had only 18 chances out of 100 in its favor, and the chance that a mental age was within three years correct was 95.7 to 4.3.

In the light of this information it may be of interest to present the probable error of a mental age derived from the application of the block-design tests.

The data of the table on the opposite page were secured from the figures of Table LXX (p. 176).

Although the median quartile deviation is approximately two months lower than the average (arith. mean) quartile deviation, the latter, although less favorable to the block-design tests, will be utilized in this analysis.

The data of Table LXXXI indicate that the probable error of a block-design mental age in terms of Stanford-Binet mental age is about 16 months. In other words, the chances are that a mental age based upon block-design performance will not be incorrect more than 16

TABLE LXXXI

MEDIAN BINET AGES FOR EACH BLOCK-DESIGN AGE

	Block-design Age* (Yrs.)															
	Below 6	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
No. of Cases	43	40	29	30	38	33	27	28	22	20	18	12	14	3	4	5
Median Binet Age* . .	64	90	98	108	105	120	129	133	138	144	149	181	186	181	191	202
Q_1 . . .	56	78	88	93	96	104	108	121	130	126	137	138	177	160	181	195
Q_3 . . .	73	102	109	120	123	136	138	150	157	164	165	217	209	204	229	207
Quartile Deviation . .	8.5	12.0	10.5	13.5	13.5	16.0	15.0	14.5	13.5	19.0	14.0	39.5	16.0	22.0	24.0	6.0

* The figures in the body of the table are in terms of months.

Average Quartile Deviation . . 16.1 mos.
Median Quartile Deviation . . 14.25 mos.

months in at least 50 per cent of the cases examined. The table of ratios follows:

Deviation	*Per Cent of Cases*[1]
12 mos.	61
16 mos.	50
18 mos.	45
24 mos.	31
30 mos.	20
36 mos.	13
42 mos.	8
48 mos.	4

$$r = + .84$$

A difference in mental age as great as four years may be expected in only 4 cases out of 100, a difference of three or more years in only 13 cases out of 100, and a difference of two years or less may be expected in 69 cases out of 100.

Based upon this criterion alone the block-design test appears to have about 56 per cent the diagnostic power possessed by the Binet vocabulary.

(9) *The Segregation of the Feeble-minded*

Although at the present time the writer has no evidence to adduce other than his experience coupled with a certain insight into the general condition of feeble-mindedness, it seems reasonable to assume that feeble-mindedness is not an arbitrary psychological and

[1] These results have been derived through the use of Sheppard's table of values of the normal probability integral modified for interpolated terms and presented in Table XC in Appendix III. Because of the interpolations this table may be found to possess greater utility than Thorndike's "Mental and Social Measurements" (2d ed., p. 200) or Buckingham's "Spelling Ability," p. 35 and p. 116.

statistical designation which fluctuates and changes as do our moral standards. After years of effort in this field one cannot help but be led to the conclusion that mental feebleness is merely *one* of the *surface* indications of general physiological incompetence. This notion is expressed as follows by Cabot:

> "Our best social diagnoses, such as *idiocy* or *feeble-mindedness*, do not refer to the mind only. They refer to the body just as much. Feeble-mindedness is a statement about the child's body, his brain, his voracious appetite, the diseases to which he is likely to succumb, his extraordinary susceptibility to cold, and his poor chances of growing up. One says a great deal about the physical side of a child, as soon as one pronounces the word 'feeble-minded.'"[1]

It is just this aspect of the general problem which has been overlooked by those whose only basis for the term "feeble-mindedness" has been the "normal curve of distribution." Mental deficiency is a general, disordered, physiological condition, which transcends such simplicity of definition or understanding.

It has frequently been argued that because the results of an intelligence test yield, what appears to the eye, a normal distribution curve, this in itself is complete proof that human intelligence is a perfectly-graded, homogeneous entity. The writer, at this time, merely wishes to call to the reader's attention the important discussion by Boring[2] in which he points to the fact that

[1] Richard C. Cabot: "Social Work," (Boston: Houghton, Mifflin Co., 1919; 188 p.), pp. 107–108.

[2] "The Logic of the Normal Law of Error in Mental Measurement," *Amer. J. of Psychol.*, 1920; 31: 1–33. See especially pp. 12, 13, and 15.

whereas Merriman's frequently quoted table of shots aimed at a target is generally utilized to illustrate the functioning of the normal law to errors, nevertheless "Pearson showed that the distribution of shots at a target, given by Merriman, is best represented by the sum of normal curves." For the same reason there is room for belief that possibly the curve for human intelligence, which in reality is not "normal" but skewed, may be the composite of two (perhaps more) separate curves, perhaps "normal," perhaps not, each representing measurements of somewhat distinctively separate groups.

Any test devised to measure intelligence has a specific bearing upon the general problem of the clinical differentiation of the feeble-minded. The present test has been purposely devised to be of especial assistance in this general direction. After wrestling with various problems of clinical psychology for a number of years, one of the pressing outstanding questions appeared to be that of the borderzone case. A number of mental examiners apparently are content with designating a child "borderzone intelligence," and allowing the matter to rest there. This is especially true of those who have been given to a rather arbitrary ultilization of the Binet scale, chopping off cases into groups on the basis of I. Q. alone. So, in spite of the fact that any determined I. Q. may be in error 2 or more points in 50 per cent of the cases, if Johnny Smith is 79 I. Q. he is a "borderzone case," but if it happens to be 81, he is just dull normal! Clinical psychology, to gain substantial recognition, must take itself more seriously! Binet I. Q. is merely

one of the minutiæ of an adequate diagnostic syllabus. And a final summation of information from the psychological, educational, social, medical, vocational, and biological fields is considerably more than can be epitomized in a brief label of one or two words.

The writer is making a plea, not for the abandonment of intelligence classifications as such, but rather that the *psycho-clinician* dictate his classification, and not allow the classification to master and dictate to him. There are frequent occasions when established classifications fail to fit a specific case. That is the time when the efficient examiner manifests his mettle and his intelligence by freeing himself from any previous systematized procedure and analysis, and arrives at the truth by other channels and other criteria.

Looking at the matter from a social and practical point of view, those who engage the services of trained clinical psychologists do not look for any half-baked theories or any speculative conclusions about certain individuals or families who are distinct social problems. They require definite, practicable, intelligent recommendations, based upon a thorough and scientific analysis of the factors involved. Shall this child be sent to an institution for the feeble-minded? Can Mary, aged nineteen, be held responsible for her acts? What shall we do with Johnny, the recidivist, father insane, mother out working? Shall the Jones family, consisting of step-father, mother, and eight children, be broken up? Ought we to sterilize William? Shall we place Mrs. Smith as a domestic? Shall Tom be sent to the insane asylum or the penitentiary? What can we do for the Thompson

family, a typical instance of continued dependency running over a considerable period of years? And so on, *ad infinitum*. All along the line the life, happiness, and futures of our fellow human beings are being placed in the balance, and we aid in tipping the balance, up or down. Terms such as "borderzone intelligence" based upon the Binet test mean little if anything in writing out a prescription for action necessary to be taken in each of the above-mentioned instances. The social worker who is to handle a case wishes to know whether there is reason to hope for a moderate degree of success at social rehabilitation, or whether hope of success is unthinkable in the light of existing handicaps and social forces. To him a person is normal or not, well or not, dead or alive. One who is feeble-minded should not be handled as normal, he who is normal should not be handled as feeble-minded. For the feeble-minded the road-directions point one way, for those who are normal they point in the opposite. There exists no borderzone.

Whatever theories we may hold regarding the development, the distribution, the varieties and classifications of intelligence, a person presented for examination is or is not feeble-minded. Our determination of this fact may take on various degrees of expertness. In standardizing the block-design tests the writer was motivated by a desire to devise an additional tool which might assist in an intelligent determination of whether a person was or was not feeble-minded.[1] Among ourselves we may

[1] As used here, the term "feeble-minded" is understood to be based upon a thorough analysis of all phases of a person's past and present history taken from all diagnostic angles, and not upon the Binet test alone.

create all sorts of fantastic notions and classifications about which we may pleasantly wrangle, such for example as the "borderzone group," some of us believing that that group represents the overlap between normals and feeble-minded, and others of us maintaining that that group is a distinct, clear-cut entity, — but to the practical worker engaged in curing the ills of humanity these "fantastic notions" mean next to nothing.

What has been said may appear a rather unique and veiled return to an earlier contention regarding the "borderline case"[1] which some may have expected the writer to have abandoned after some hoped-for increase in maturity and knowledge! The chief purpose, however, in the present instance was to emphasize the importance of "differentiability" which a good test of intelligence should possess. For a very adequate discussion of the borderzone question the reader is referred to Miner's monograph "Deficiency and Delinquency."[2]

Five cases have been selected from the writer's most recent clinical experience to illustrate the assistance which the block-design tests may render. This material was written on August 3, 1921, and the five cases listed below represent the last five consecutive examinations in which the block designs were used.

[1] Kohs, S. C.: The Practicability of the Binet Scale and the Question of the Borderline Case. Bull. No. 3, 1915, Research Dept. Chic. House of Correction, p. 23.

[2] Miner, J. B.: "Deficiency and Delinquency" (Baltimore, Warwick & York, 1918; 335 p.).

KEY TO ILLUSTRATIVE CASES

Name	*Tests Used*	*Date Examined*
1. Irving T.....	(*a*) Illinois Examination II, Form 1 (*b*) Stanford-Binet (*c*) Kohs Block-designs	July 28, 1921
2. Edna U.....	(*a*) Stanford-Binet (*b*) Kohs Ethical Discrimination [1] (*c*) Kohs Block-designs	July 27, 1921
3. Chas. V.....	(*a*) Stanford-Binet (*b*) Kohs Block-designs	July 21, 1921
4. Frank W....	(*a*) Stanford-Binet (by Miss K. Donald) (*b*) Kohs Block-designs	July 14, 1921
5. Kenneth X..	(*a*) Stanford-Binet (*b*) Kohs Block-designs	July 13, 1921

Case No. 1. At the time of the examination Irving was 17 years 10 months old. Born in Minnesota, of American-born parents, German-English descent, reached the seventh grade, now plumber's helper earning $5 a day, has a small bank savings. Father died little over two years ago, was a lumber piler. Three other children, older, one boy, two girls. Mother has no influence over Irving. Lives in rented two-story house. Home fairly furnished, clean, and orderly. Irving well clothed. History of delinquency may be tabulated as follows:

[1] Kohs, S. C.: "An Ethical Discrimination Test." *J. of Delinquency, 1922, Vol. 7, pp. 1–15.* Copies of this test with complete instructions and scoring material may be secured from the C. H. Stoelting Co., 3037 Carroll Ave., Chicago, Illinois

Date	*Delinquency*	*Disposition*
Oct. 13, 1917	Stole sample case containing gums, candies, cookies from auto. Sample case found. Contents gone. Implicated with three other boys.	Cautioned and released.
Sept. 19, 1918	Stole box of crackers from —— Cracker Co. Irving admitted theft. Claimed he was hungry. Box unopened when found.	Ordered to return box, apologize and work to pay damages. Cracker Company satisfied with apology after box was returned. Warned that if he appeared in Court again might be sent to one of the institutions.
June 9, 1919	Maliciously abusing another boy.	Advised to desist and reminded of the State Training School.
July 26, 1919	Larceny of automobile. Captured after running machine into telephone pole and front of a grocery.	Placed on probation with a Big Brother who arranged for boy's employment. Damages to be paid.
Jan. 7, 1920	Stole two chickens. Three other boys implicated.	Severely lectured.
Nov. 7, 1920	"Shooting craps" on street.	Attention of parent called to this matter.
Mar. 10, 1921	School attendance officer reports Irving will not attend evening school. Father dead, mother can do nothing with boy.	No specific action taken.
July 27, 1921	Larceny of one auto on July 17, confessed to another on July 4. On July 18 stripped accessories from two other machines.	Disposition now pending.

On the Illinois Examination II, which the writer has adapted for individual testing, Irving obtained a total score for the general intelligence tests (7) of 67 points. This is equivalent to a mental age of 11 years 0 months. His I. Q. comes to 69 per cent. On the Stanford-Binet, with a basal year of 9, he obtains an intelligence rating of 12 years 4 months, giving him an I. Q. of 77 per cent. On the block-design test, seven out of thirteen designs attempted are successfully completed, yielding a score of 35 points which is equivalent to a mental age of 10 years 11 months. His I. Q. here is 68 per cent, within one point of the Illinois Examination.

Irving, on the basis of the Binet test, is a borderzone case. On the two other tests he falls below the 70 per cent dividing line. Taking a composite, or an average of the three estimates, his I. Q. becomes 71, still borderzone, which was the intelligence classification finally set down. The complete social analysis of the case is in the process of being made, but the psychological report, at the present stage reads that "if continued social incapacity is manifested in the future, this boy may be classified as feeble-minded and as such committed to the State Institution for the Feeble-minded."

Case No. 2. Edna is a nice-appearing girl, 15 years and 11 months old at the time of examination, and rather over-developed physically for her age. Born in Oregon of American-born parents, Irish-English descent. Reached the seventh grade. Mother died of tuberculosis soon after Edna was born. Father thought dead, but only recently traced to a small town in Oregon where he is employed as a mill-worker (general labor). Edna has

a brother about four years older, about whom little is known. She first became notorious about a year ago when charges of contributing to the delinquency of a minor were brought against the principal of a small town school which she was attending. The principal was almost lynched at the hands of the inhabitants, although acquitted of the charges by a jury. After this episode she was brought to Portland and placed with an aunt who is rather irresponsible. Excerpts of the report from the Woman's Protective Division read as follows:

"Sat. July 16, Edna was sent to a grocery store at —— Sts., to make some purchases. D—— waited on her. She loitered in the store looking at a book and finally D—— told her she owed him 10 cents and asked her to come into the back room. She went in and D—— had relations with her. She says she went to the store between 1 and 2 and left between 2 and 3.

"She dared not go home so started to walk. On —— st. she was overtaken by Ed —— who was driving an automobile. (Ed —— is a juvenile who has previously been before the Court for immorality.) He asked her where she was going and she replied to Q—— (a small town approximately 20 miles distant). He offered to take her there; she got into the machine with him, he drove her to Q—— out —— Road and had relations with her in the woods. Then took her back to Q—— at about 10 P.M.

"She stayed alone at —— Hotel, room 6, registered as —— —— (fictitious name) and paid 75 cents for her room. She says she had a little money when she left home.

"Ed—— gave her his name and address when he left her. In the morning she started to walk back to Portland, had no supper or breakfast. She walked all the

way back to Portland to ——, arriving there about 3 or 4 Sunday P.M.

"George —— saw her standing on the street corner across from —— St. He and Ed —— took her in a Ford for a ride, went out past —— Road, both boys had relations with her, first Ed, then George. George had relations with her twice at this time. She thinks she was out with the boys about two hours . . ." And so the story unravels, involving more boys the farther it goes. The girl was not actively in search of these experiences but seemed a rather obvious mark to the unscrupulous. When this information reached the press it appeared under the following headlines:

"GIRL, 15, FOUND IN SHACK, 20 men suspects arrested, but only 5 are held — police are of opinion runaway child is weak-minded — physician takes charge."

After a short period of observation Mrs. Lola G. Baldwin, head of the Woman's Protective Division, declared the girl utterly irresponsible.

By the Stanford-Binet test, with a basal year of 8 years, Edna manifested an intelligence equal to 11 years 5 months. With a life age of 15 years 11 months her I. Q. is 72 per cent — borderzone. Given the block designs she succeeded in passing only one, the first, of the six designs attempted. Her score on this test amounted to 1 point, this being equivalent to a mental age of 5 years 7 months. This indeed, is far from the girl's actual intelligence level, but it possesses diagnostic significance, especially in relation to our "borderzone" problem. On the Ethical Discrimination Test the percentage efficiency in each exercise was as follows:

1. Social Relations	90%
2. Moral Judgment	80%
3. Proverbs	60%
4. Definitions of Moral Terms	53%
5. Conduct Evaluation	50%
6. Moral Problems	90%
Average	71%

On the basis of tentative norms this average of 71 per cent is equivalent to the performance of about a 12½ year-old.

In the light of this history, her motives, and her general record, the girl was declared feeble-minded, unable because of marked retardation in intelligence to manage herself and her affairs with the ordinary prudence common to other girls of her life age. The further disposition of the case is now pending.

Case No. 3. Charles is the duller-looking one of the two boys left with a maternal aunt after the father died five years ago, having been killed by a railroad train, and the mother had been committed to State Institution for the Feeble-minded two and a half years ago. Mrs. M——brought Charles to court one day (Dec. 2, 1920), claiming the boy was feeble-minded just like his mother, that she was willing to keep the other boy, but that Charles was a hopeless case and consequently the state should undertake the care of the child. At the time of the first examination Charles was 7 years 11 months and tested, by the Stanford-Binet, showed an intelligence of 6 years 6 months, his I. Q. being 82 per cent, and classifiable as dull-normal. Although a petition in feeble-mindedness had been duly filed in this case, this was canceled and the boy turned over to the Boys' and Girls'

Aid Society for observation, education and treatment. Charles was born in Portland of German-American parentage, the father only being foreign-born. The child's report card for three months showed "good" marks in all subjects except "application" for the second month. Charles is affected with ptosis, giving one a poor impression of his intelligence. On April 12, 1921, we unearthed the fact that a quack "psycho-analyst," had "examined" the boy back in 1918, had adjudged him feeble-minded, a vacancy for the boy had been made at the feeble-minded institution, and in 1919 when the boy was about to be sent, his aunt experienced a change of heart. She claims Charles is unreliable, can remember nothing, frequently fails to come home from school or after being sent upon errands, and that he has been two years in the first grade. There is a sister of Charles' eleven years old staying with grandparents in eastern Oregon who are very poor. When Mrs. M's home was visited it was found to be a four-room affair, rather neat. Mr. M was found in bed where he has been confined for many years with articular rheumatism. The sources of income of the M family are distributed as follows: "$20 a month from Mr. M's lodge, sick-benefit; $30 from the county for the care of Charles and his brother, and the money paid by three men boarders. Of this money $100 must be paid every three months in payment for the purchase of this home. Mrs. M has three daughters, aged 18, 16 and 8. The two eldest are in high school and must sleep at a neighbor's for lack of room. The boarders occupy a tent in the yard. After Charles was at the Boys' and Girls' Aid Society for four months,

he was reported by the Superintendent as being feeble-minded and unplaceable. Another mental examination was, therefore, arranged. On July 21, 1921, Charles was again examined. He was now 8 years 6 months and on the Stanford-Binet obtained a mental age of 7 years 0 months. Again his I. Q. was 82, and was again classifiable as dull normal. It was then decided to give him the block-design tests. Although on the basis of life age he should have secured a score of 14 points, and on the basis of mental age his block-design score should have been at least 6 points, he failed to perform any of the designs correctly, in spite of continued explanations and demonstrations. His intelligence age here was below 5 years 3 months, yielding an I. Q. under 62. However, he was sent back to the Aid Society for further observation and treatment. In the meantime his intelligence has been set down as dull normal, but possibly potentially feeble-minded.

Case 4. Frank W. first came to the attention of the court when a neighbor reported on December 14, 1918, that Frank, 11, and his sister of 4 had been left with a maternal grandmother, a widow, the mother having died recently, and the father having deserted, his whereabouts not being known. The family also was reported to the local Associated Charities as being in destitute circumstances. It was learned later that Frank's father was rather nomadic in tendency, and in 1914 had deserted the family for good after the second child had been born. On September 2, 1919, it was reported to us that the grandmother was abusing and unmercifully beating Frank. On September 8, the grandmother was

visited. She claimed Frank was unruly and disobedient, that she *does* punish him, there being no other way of getting along with him; denied most emphatically that she had ever punished him to an extent that might be deemed cruel or excessive. She was found living in a rented home, her income being derived from little sewings and from keeping a boarder. For some intangible reasons she has been ostracised by her neighbors. Although Frank did not appear quite normal to the visitor, his sister seemed a happy, healthy child, neat, apparently well cared for, and devoted to her grandmother. At the time of this visit Frank was attending the regular schools, was in the 4th grade, and the teachers gave a good report of his conduct in school. Four months later we find Frank in the Etna School (ungraded) with a petition for feeble-mindedness filed by the Public Welfare Bureau. He was examined July 14, 1921. His life age was 13 years 4 months, and on the Stanford-Binet he secured a mental age ranking of 8 years 9 months. His I. Q. was 66. On the block-design test, of six designs attempted two were successfully completed, scoring 3 points on the first design and 4 points on the third making a total of 7 points. He was utterly lost beyond the third design. His mental age ranking here was 7 years 3 months, with an I. Q. of 55 per cent. The grandmother admitted that the boy was feeble-minded but was prone to argue about the causation. Disposition of the case is now pending.

Case 5. Kenneth X. first came to our attention June 20, 1921, when Miss Ida M. Manley, principal of the Etna School (ungraded) filed a petition in feeble-minded-

ness. This was accompanied by a memorandum which contained the following information. Mother until recently a janitress in one of the schools, father died two years ago, said to have had locomotor ataxia. Kenneth has been in special class for three years (third grade). Has habit of taking everything loose about the school. Is exceedingly clever in making up stories to cover his guilt. Repeated scoldings, warnings, and punishments have no effect. Kenneth's mother has had three other children, one died because of premature birth, one was born paralyzed, and the other died at six years of age of a "natural disease," as the mother put it. With a basal year of six Kenneth obtained a mental age of 8 years 10 months. His life age at the time of examination was 11 years 6 months, giving him an I. Q. of 77 per cent, classifiable as borderzone. On the block-design tests, of ten designs attempted only three were successfully completed, and after design six he appeared hopelessly baffled. His total score was 11, 1 point on design 1, 5 on design 4, and 5 on 6. This performance is equivalent to an intelligence age of 8 years 0 months, yielding an I. Q. of 70 per cent. Other elements in the situation made it appear more reasonable to hold up proceedings on the commitment for another year, pending further observation and treatment.

The cases cited have not been specially selected, but have been taken in consecutive order. The limitations of space have made it necessary to omit much social and psychological information which is pertinent. The chief object has been to indicate wherein the block-design tests have been of assistance; consequently their

importance may appear exaggerated from the bare perusal of the material presented. It is expected that the reader will understand the correct proportion of emphasis placed upon block-design performance.

In conclusion, the block-design tests, even though it should be granted that they fail to measure intelligence as efficiently as other standard methods, nevertheless are found valid because they assist in the differentiation of the borderzone group, some of whom are normal and some feeble-minded.

(10) *Percentage Ratios of Agreement and Disagreement with Established Standards of Intelligence*

In this section an analysis will be attempted to indicate in quantitative terms the actual amount of divergence between the mental age and I. Q. estimates of the block-design test and the Binet.

(*a*) *The Method of Direct Comparison:* Taking Table LXX (p. 176) and analyzing it for agreement and divergence of mental ages, the table on p. 225 is obtained.

It is apparent that the block-design test is inclined to give somewhat higher mental ages, in general, than the Binet. Interpreted in a slightly different fashion, the following tabular analysis is obtained.

Exact correspondence	and	within		1 yr.	= 53.8%
"	"	"	"	2 yrs.	= 75.9%
"	"	"	"	3 yrs.	= 89.1%
"	"	"	"	4 yrs.	= 93.7%
"	"	"	"	5 yrs.	= 98.4%

In other words, in practically 90 per cent of the cases examined the two mental ages were the same or showed no difference greater than 3 years in either direction.

TABLE LXXXII

MENTAL AGE DIVERGENCES

Per Cent Exact Correspondence	Binet Age Per Cent Higher							Binet Age Per Cent Lower						
	1 Yr.	2 Yr.	3 Yr.	4 Yr.	5 Yr.	6 Yr.	7 Yr.	1 Yr.	2 Yr.	3 Yr.	4 Yr.	5 Yr.	6 Yr.	
20.5 . .	18.0	9.3	5.5	1.6	2.5	.5	.5	15.3	12.8	7.7	3.0	2.2	.5	
	Total — 37.9%							Total — 41.5%						Total 99.9%

This substantially corroborates the findings discussed under the heading, "The Probable Error of a Block-design Mental Age."

In order to appreciate the true significance of this divergence the discrepancies between two Binet tests of the same individual will be stated.

The following has been obtained from Terman's table[1] showing the agreement between the earlier and later tests of 428 children by the Stanford-Binet. (Correlation, $r = + .93$.)

TABLE LXXXIII

I. Q. DIVERGENCES

Per Cent Exact Correspondence	Second I. Q. Points Higher (%)							Second I. Q. Points Lower (%)									
	5	10	15	20	25	30	35	5	10	15	20	25	30	35	40	45	
30.0 . .	25.0	11.8	3.0	1.9	.7		.2	14.2	8.6	2.6	.9	.2				.2	
	Total — 42.6%							Total — 26.7%									Total 99.3%

Although the basis of comparison is not entirely equivalent, some idea may be obtained of the reliability of the block-design tests if compared with the discrepancies found upon measuring the same individual with the same measuring scale on two different occasions, the I. Q. in this instance being assumed to remain fairly constant. (Although the following observation is somewhat aside, it may be of interest to call attention to the fact that the positive discrepancies of I. Q.'s obtained from the second Binet were more than one and one-half

[1] L. M. Terman: "The Intelligence of School Children," p. 143 (Table 26).

times as frequent as the lower ones.[1] It will be observed that exact correspondence occurs in 30 per cent of the cases, and the other deviations may be tabulated as follows:

Exact correspondence or within 5 I. Q. = 69.2%
" " " " 10 I. Q. = 89.6%
" " " " 15 I. Q. = 95.2%
" " " " 20 I. Q. = 98.0%
" " " " 25 I. Q. = 98.9%

It should be borne in mind, however, that at different life ages, I. Q. changes of the above magnitudes will possess different equivalent age values.

These figures serve as a fairer basis of comparison in determining the discrepancies between I. Q. as determined by the Binet and the block-design tests. (Again the comparison is not a fair one in view of the existing controversy over the constancy of the I. Q. The writer has been unable to find material which would be exactly comparable.)

Taking the date of Table LXXII (p. 183) the following agreement and divergence of I. Q. are found:

TABLE LXXXIV

I. Q. Divergences

Per Cent Exact Correspondence[2]	Binet I. Q. Points Higher					Binet I. Q. Points Lower					
	10	20	30	40	50	10	20	30	40	50	
23.5 . .	20.8	9.0	4.6	2.5	.3	18.0	11.8	5.5	2.7	1.4	
	Total — 37.2%					Total — 39.4%					Total 100.1%

[2] Frequency interval is 10 I. Q.

[1] For additional interpretations of this condition see E. A. Doll: "The Growth of Intelligence," *Psychological Monographs*, Vol. 29, No. 2, 1921, 130 p., especially Chap. III.

Or presented in another form:

Exact correspondence or within	10 I. Q.	= 62.3%
" " " "	20 I. Q.	= 83.1%
" " " "	30 I. Q.	= 93.2%
" " " "	40 I. Q.	= 98.4%

A better balance between the per cent of negative and positive divergencies is observable here, and also a greater spread of variability than is the case with I. Q.'s derived from consecutive Binet testings.

(*b*) *The Method of Critical Ninths:* A determination of discrepancy may be obtained by an analysis of a different sort. Every table presenting the relation between the frequencies of two variables may be divided into nine parts, Q_1 and Q_3 for both serving as the dividing lines. Q_1 and Q_3 divide any distribution into three portions low, middle, and high. Consequently nine groups are possible, thus:

Variable (1)

Variable (2)	Low (Q_1)	Middle (Q_1–Q_3)	High (Q_3)
High (Q_3)	L_1 H_2 (*critical cases*)	M_1 H_2	H_1 H_2
Middle (Q_1–Q_3)	L_1 M_2	M_1 M_2	H_1 M_2
Low (Q_1)	L_1 L_2	M_1 L_2	H_1 L_2 (*critical cases*)

For two tests which supposedly are expected to measure the same trait, L_1 H_2 and H_1 L_2 may be regarded as the critical sectors in which as small a proportion as

possible of cases should be found. Every case found within these limits is a clear argument against the validity of a newly devised instrument, when measured against one which has already been established (assuming that the already-established instrument is absolutely reliable).

Applying this criterion of validity to the data of Table LXX, the following percentages are obtained:

Variable (1) Binet Age
Variable (2) Block-design Age

Critical Cases

$L_1\ H_2 = 0\ (0\%)$
$H_1\ L_2 = 2\ (.5\%)$
Total = *0.5%*

Correspondence

$L_1\ L_2 = 61\ (16.7\%)$
$M_1\ M_2 = 129\ (35.3\%)$
$H_1\ H_2 = 71\ (19.4\%)$
Total = *71.4%*

In order to make these values comprehensible in terms of other correlations, the percentage of critical cases for the Binet and Vocabulary[1] and for Army Alpha and Stanford-Binet Age[2] will be presented:

Variable (1) Binet Vocabulary
Variable (2) (631 cases) Binet Age

Critical Cases

$L_1\ H_2 = 0$
$H_1\ L_2 = 0$
Total = *0%*

[1] L. M. Terman: "The Intelligence of School Children," p. 310.

[2] R. M. Yerkes: "Psychological Examining in the Army," (Memoirs, Nat. Acad. of Sciences, vol. 15, 1921, 890 p.), p. 621.

Correspondence

$L_1\ L_2 = 145\ (23.0\%)$
$M_1\ M_2 = 244\ (38.7\%)$
$H_1\ H_2 = 113\ (17.9\%)$
Total = *79.6%*

Variable (1) Stanford-Binet Mental Age
Variable (2) Alpha Raw Score (653 cases)

Critical Cases

$L_1\ H_2 = 0$
$H_1\ L_2 = 1\ (.15\%)$
Total = *0.15%*

Correspondence

$L_1\ L_2 = 104\ (15.9\%)$
$M_1\ M_2 = 273\ (41.8\%)$
$H_1\ H_2 = 99\ (15.2\%)$
Total = *72.9%*

It is apparent that a close correspondence exists between the evaluations of intelligence (in terms of Binet mental age) on the part of the vocabulary test, the army alpha, and the block designs.

Finally, the same analysis is presented for Binet I. Q. and Block-design I. Q. (data of Table LXXII, p. 183).

Variable (1) Binet I. Q.
Variable (2) Block-design I. Q.

Critical Cases

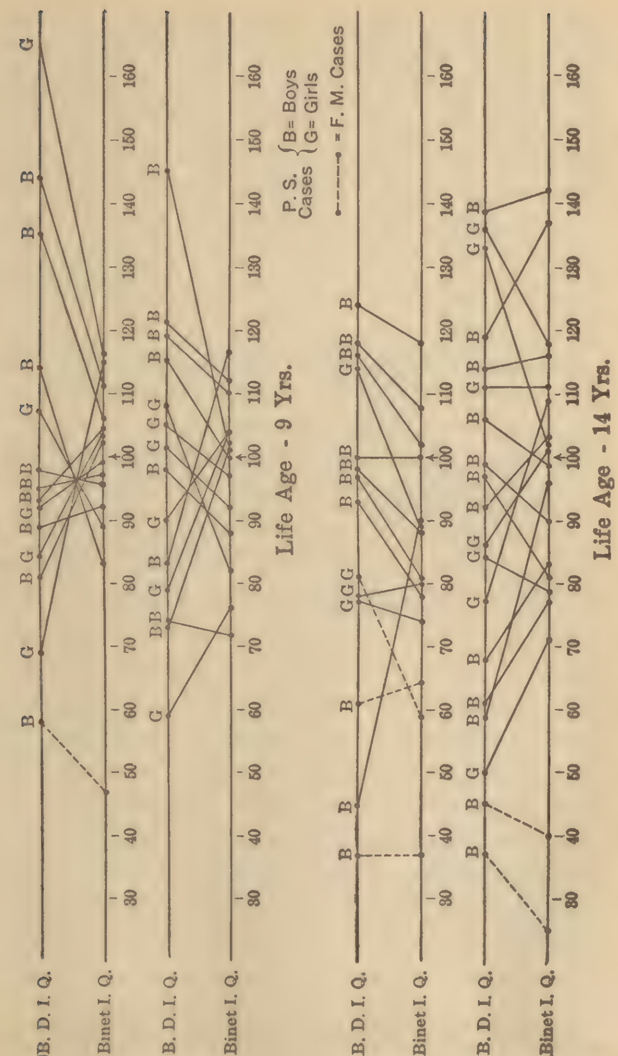
$L_1\ H_2 = 2\ (.5\%)$
$H_1\ L_2 = 0$
Total = *0.5%*

Correspondence

$L_1\ L_2 = 73\ (19.9\%)$
$M_1\ M_2 = 131\ (35.8\%)$
$H_1\ H_2 = 51\ (13.9\%)$
Total = *69.6%*

GRAPH 43

I. Q. Discrepancies



(*c*) *Graphic Analysis:* For a graphic presentation of discrepancies the life ages of 14 and 9 have been selected — age 14 because this life age had the largest single number of feeble-minded below age 17 (see Table I) and age 9 purely at random. An analysis of Graph 43 indicates the existence of a number of conditions:

(1) A mild process of compensation is operative, some who have low I. Q. on the Binet, obtain high I. Q. on the block designs and vice versa (especially true of the 9 year-olds);

(2) Low Binet I. Q. are often coupled with still lower B. D. I. Q., and high Binet I. Q. are often found in conjunction with still higher B. D. I. Q.

(3) The larger proportion, however, manifest little, if any, discrepancy between the two I. Q.

It will be observed that the range of I. Q. in the block-design tests is considerably increased. This has its disadvantages, but from a clinical point of view this characteristic is a desirable one, because of its bearing upon the differentiation of the inferior and the superior.

Summarizing this discussion of discrepancy, it may be stated that although the results of the block-design test do not tally exactly with those of other measures of intelligence, nevertheless, agreement is close enough to grant the block designs a fair degree of validity, the instances of disagreement being in themselves sufficiently crucial to require a psychological analysis apart from the tests themselves.

This concludes the discussion of validity. A number of criteria have been utilized, each differing in weight and

in general importance. In reality the only effective acid test is the pragmatic one. If the tests will be found to work, that is all that the most critical ones of us demand.

The array of evidence just presented seems to indicate not only that the tests measure intelligence, but that this is accomplished with a fair degree of accuracy. On the other hand, one should bear in mind Stern's caution:[1]

> "Psychological tests *must not be overestimated*, as if they were complete and automatically operative measures of mind. At most, they are the psychographic minimum that gives us a first orientation concerning individuals about whom nothing else is known, and they are of service to complement and to render comparable and objectively gradable other observations — psychological, pedagogical, medical — not to replace these."

3. DIAGNOSTIC VALUE OF THE TESTS

It has been intimated a number of times by writers of varying degrees of expertness and reliability that an estimate of intelligence based upon the use of a single measuring instrument is inclined to be less accurate than one based upon the application of a number of such diagnostic tools. In fact, a frequent failure upon the part of numerous psycho-clinicists is to disregard other important criteria for the proper diagnosis of intelligence or of mental deficiency and to rest content with the results of a single test, somewhat mechanically evaluated. This is especially true of feeble-mindedness. No diagnosis of this condition is adequate which has not included

[1] "The Psychological Methods of Testing Intelligence" (Warwick and York, 1914), p. 12.

weighing the evidence in other fields than the psychological. Thus a verdict of feeble-mindedness which is not supported by evidence of vocational impotence, of educational failure, of social incompetence, of biological taint, is, except in rare and exceptional instances, quite weak and untenable.

The issue also has been raised whether we have not overemphasized the verbalistic tests with the inevitable injustice to those who may be laboring under a language handicap. Without desiring at this point to enter any evidence for or against the reasonableness of this contention, other than that already presented in this monograph, it does seem wise to utilize tests which measure intelligence through "performance" as well as through "language." The combined results of an inquiry in both these directions seem to indicate a closer approximation of what is the *actual* mental age of a subject, than if either is used alone.

On the other hand, it is plainly evident that no "performance" test yet devised measures intelligence as adequately or with as high a coefficient of efficiency as the Binet or two or three other "language" tests. Of course, the standards upon which our judgments of efficiency have been based may be unsatisfactory, e.g., school progress, teachers' estimates, school marks, etc. At any rate, looking at the matter from a practical point of view, few, if any, clinical psychologists would be free to abandon our so-called "language" intelligence tests for the exclusive use of "performance" tests in classifying children and in making diagnoses of mental deficiency. At present, performance tests are of chief importance as

corroborative and supplementary sources of information. This probably will hold true for some time to come.

4. SERVICEABILITY

In his "Stanford Revision of the Binet-Simon Scale" (Warwick and York, 1917) Terman states,[1] that "to be widely serviceable a test should demand only the simplest material or apparatus, should require at most but a few minutes of time, and should lend itself well to uniformity of procedure in application and scoring." The writer has attempted to satisfy these demands in standardizing the block-design tests. Those who utilize the tests will find after a little practice that there can be but little variation in the findings of two examiners, and that the only chance for difference is in the recording of the number of moves made.

The special value of the block-design tests lies in the fact that valid results may be obtained independently of the "language factor." Neither deafness nor lack of language-understanding should be disqualifications in the proper performance of the test. The block designs may therefore be utilized in the study of racial differences, in determining the mental capacities of the deaf and of those suffering from various other language handicaps.

As regards the borderzone problem, it appears that this test will aid in a better differentiation of the group of cases falling in this category. The writer maintains that feeble-mindedness is *not* an arbitrary statistical designation, but is rather a clearly demarked physiological entity quite distinct from normality, statistical psychologists

[1] P. 150.

notwithstanding. Years of experience with this type of defect has fixed the notion in the writer's mind that feeble-mindedness is indicative not only of mental mal-functioning, but also of physiological mal-functioning possibly of endocrine character. The results of further research, however, can be the only tests of the truth of one's statements at this time.

CHAPTER VI

Supplemental Observations

1. Average Mental Age of Adults
2. Increasing Divergence versus Parallel Progress of Mental Development
3. A Performance Intelligence Scale

1. AVERAGE MENTAL AGE OF ADULTS

Of importance in interpreting the results of this newly devised mental test is the recently raised question regarding the average mental age of adults. The current practice of regarding 16 years as the age when intelligence-maturity is attained, and the age characteristic of the American population at large, has been adhered to in the standardization of the block-design tests and in the interpretation of scores derived therefrom.

The results of army testing have thrown some question upon the tenability of this standard, and although previous to the testing in the army criticism of the 16-year criterion emanated from certain quarters,[1] no great importance was attached to these attacks, clinical psychologists continuing the practice of regarding 16 years as the intelligence of the average adult. The results of army testing, due to the overwhelmingly large numbers examined, are likely to receive greater recognition and

[1] See controversy between J. E. W. Wallin and S. C. Kohs in *Journal of Amer. Institute of Criminal Law and Criminology*, 1916, vol. 6 and 7.

may be regarded by some as unalterably conclusive. The army tests indicated that the average mental age of adults, especially utilizing the figures of Group X, a supposedly random sampling of 653 men, is 13¼ years.[1]

The attempt will here be made to present reasons why we should exercise caution in accepting this level as average, advising instead the continued use of the 16-year standard until further experimentation has been carried on to decide the issue finally. The reasons for questioning the 13¼ estimate will be listed *seriatim*, and have not been arranged in order of importance but rather as they suggested themselves to the writer.

(1) It is advantageous in measuring intelligence to utilize a tool which will not present adaptive difficulties for the person examined. Thus familiarity, previous experience, specialized imagery, the peculiar mechanics of the test itself may make it easier for some to obtain satisfactory scores to whom the technique of the test is adapted, and, on the other hand, more difficult for those to whom the test procedure and its attack are strange and unfamiliar. The army test, together with the conditions under which it was taken, in and of itself requires "an adaptation to a new situation" which is quite complex. All subjects will not adapt themselves similarly. This condition introduces a spurious factor likely to underweight the performance of those below average in mental ability, with the inevitable result that the general average is reduced. An ideal test of intelligence *measures* "ability to adapt oneself to a new situation," but is not

[1] R. M. Yerkes: "Psychological Examining in the Army," Memoirs National Acad. of Sciences, pp. 195, 790.

itself a matter to which "adaptation" is required. In this respect the army tests may be said to fall short, a correcting formula being necessary to compensate for this "drag" upon sub-average ability, when determining the average capacity of a group examined.

(2) A matter deserving serious consideration at this time when group testing has deservedly gained such great popularity, both for analyses of intelligence level and of pedagogical achievement, is what the writer has termed the "*theory of progressive score reduction*." A tacit assumption generally made in group examining is that ability manifested in a group-test will very closely correspond to the same level of ability attained in an individual examination. Regression tables and correlation coefficients are presented as evidence bearing out this contention. A close analysis of the situation will, however, reveal the existence of this condition: Whereas a high score in a reliable group intelligence test is unquestionably indicative of that mental level, because superior scores cannot be obtained by accident or haphazard, the same does not hold true for low scores. One hardly dare state that a low score is *ipso facto* the inevitable performance of an inferior mentality. Any number of factors may have entered to reduce this score. The chances are that the lower the examinee's intelligence the greater are the number of these "reducing" factors; and *vice versa*, the higher the intelligence, the lower the number. The actual situation may be pictured graphically.

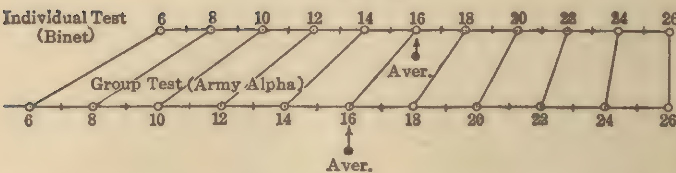
To make this graph clear, let us assume that we have applied the Binet tests to a large number of subjects and

we select eleven whose intelligences are respectively, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26.[1] The median mental age here will be observed to be 16. Let us also suppose that this large group, in which these eleven are included, have been given Army Alpha. It is the writer's contention that the lower the actual mental ability of the subject examined, to that extent will his score in the group test deviate from his true mental age. According to Graph 44, the subject testing 26 on one test will, supposedly, test exactly the same on the group test. For the other ages.

a 24 Binet will possibly measure 23½
a 20 Binet will possibly measure 18⅛
a 16 Binet will possibly measure 13¼
a 12 Binet will possibly measure 8⅛
an 8 Binet will possibly measure 3½

GRAPH 44

Age Reduction



The median performer who on the Binet obtains an intelligence rating of 16, may on the group test, because of this phenomenon of progressive score reduction, obtain an intelligence rating of barely 14. (The material

[1] The writer is aware that the Stanford-Binet cannot yield such high mental ages except through the predictive use of the I. Q. But for the purposes of this presentation, this deficiency will be overlooked.

of the graph has been made to conform with the possible condition of the figures obtained from the application of Army Alpha.) It should be noted that the data of Graph 44 would yield a correlation coefficient of + 1.00, yet it would not reveal the increasing inaccuracy of the group intelligence examination with the decrease in the mental ability which it has been devised to measure.

(3) It cannot be denied that the Army Alpha overweights reading ability. Those who have difficulty in reading, or in translating the printed word into more familiar imagery, may be expected to obtain lower scores than on an individual test like the Binet. Evidence of this was clearly revealed by the fact that going from Army Alpha to the Combined Scale did not yield the same results in terms of Combined Scale scores, as going from Army Beta to the Combined Scale. In the Beta test (performance) those possessing the mental ability had a more satisfactory opportunity to manifest it than with Alpha. So that in the case of Alpha both the dull normals and inferiors, together with those laboring under language handicaps of whatever form, tended to slide toward the lower end of the frequency distribution.

Attention, in this connection, is called to the recent experience of Whipple [1] with three pupils who obtained a mental age from one to three years lower on the National Intelligence Tests than on the Stanford-Binet. Quoting:

"These three cases strongly suggest, what would be *a priori* intelligible, that pupils who have special difficulty

[1] G. M. Whipple: "The National Intelligence Tests," *Journal of Educ. Research*, 1921, 4: 18–31.

in reading will suffer a decided reduction in mental age rating by a group intelligence test when compared with their rating by the individual and oral examination of the Binet, though it remains possible that the former rating may be the more significant in predicting school progress." (P. 31.)

(4) The assumption that the draft represented a fair cross-section of the American population is admittedly unwarranted. The frequency tables from which the average intelligence of the American citizen is deduced does not include the figures for officers of all grades. It should also be recalled that thousands of professional men, and those engaged in essential industries were exempted from service and were consequently free of the draft. If we can regard a man more intelligent because he has taken on family responsibilities and has been exempted because he is engaged in caring for those dependent upon him, then this is another very large and important group whose range of ability is not adequately represented in the total results.

(5) It is a recognized fact in psychological testing-technique that the physical conditions under which a subject is working may have a very profound effect upon the result achieved. It perhaps will not be questioned that the experimental conditions at the various draft camps were with few exceptions very far from ideal. It was recognized that the failure of the War Department to provide special buildings for psychological examinations proved a very serious handicap, and the facilities which were extended interfered considerably with the effectiveness of the work. Among

the factors which might be expected to affect the results were the following:

(*a*) Examinations followed at varying intervals after inoculation for typhoid, influenza, etc. at various camps;

(*b*) The same degree of variation was characteristic for the time of examination after arrival;

(*c*) In some camps men were examined with a full equipment of apparel including the heavy top-coat, in others men were in their shirt sleeves;

(*d*) Sitting on the floor, tailor-fashion, some on mats, some against posts or the wall for support, others without mats or back-supports, may have had some influence upon the results;

(*e*) The use of beaver-boards for hand-rests may have had a varying influence upon efficiency;

(*f*) The preëxamining conditions, physical and mental, may possibly have been a factor which should be considered;

(*g*) The light, air, comfort, and physical properties of the various buildings which served as examination centers were variable quantities which also may have some bearing upon the interpretations of the final averages.

It might be well to state before continuing, lest the writer be misunderstood, that these remarks are not made in criticism of the splendid work which was accomplished by the Psychological Division of the Army, but an attempt is being made to interpret the statistical evidence that the average mental age of adults is 13¼ years. In view of the difficulties and handicaps which were inevitable and insurmountable, we should be

cautious in carrying over this result to the field of clinical psychology, overturning standards which have been current practice for a good many years. Let us first check these averages for scientific validity.

(6) A necessary element in all group testing is "speed." Experimentation with the army tests has thus far revealed that "speed" and "intelligence" are fairly well correlated, and that doubling the time in Army Alpha, for example, does not materially change final rankings. It should be borne in mind, however, that possibly "speed" has been overweighted in the army tests. In fact it is admitted that "assuming that we are dealing with an unselected group, Table LXXIX shows that some limit between double and single time would probably be more suitable."[1] There is considerable question now regarding the relative values of "speed" versus "power" tests. There are many who insist that "speed" should be made subordinate to "power." In view of the possible overemphasis upon speed in the Army Alpha tests, together with the difficulty of discounting it accurately, perhaps we had better await further experimentation before accepting the intelligence-average as found.

(7) For a number of years it has been the practice of clinical psychologists to subdivide the feeble-minded into three groups: idiots, imbeciles, and morons. The mental age limits for these groups are: idiots, from below age one to age two; imbeciles, ages three to seven; and morons, ages eight to twelve. According to the

[1] R. M. Yerkes: "Psychological Examining in the United States Army," p. 420.

data of the army examinations, if all those testing less than twelve are designated feeble-minded, then 30.3 per cent would be so classified. This, of course, is absurd and unthinkable. We have been ready to admit that approximately the lowest two per cent of the population is mentally unfit to manage itself and its affairs with ordinary prudence. On this basis the army examinations indicated that those testing below eight amounted to 2.1 per cent. To reverse our reasoning and to contend that any one with a mental age of eight or over is mentally fit to manage himself and his affairs with ordinary prudence is to lead to another absurdity belying the facts of our daily experience.

We are therefore confronted with two horns of a dilemma, from which there seems no clear-cut escape. If it is true that 13¼ years is the average mental age of adults, and if it is true that 8 years is the line of demarcation for those who are feeble-minded, then we should be prepared to empty our feeble-minded institutions of at least one-fourth of their present populations. We should also proceed immediately to change the I. Q. classifications of intelligence. Instead of an I. Q. of 70 or 75 being indicative of feeble-mindedness (utilizing 16 years as the devisor for adults), the I. Q. of 60 should be adopted (13¼ years being the highest divisor possible). Similar revisions should be made for the other groups. The moron feeble-minded, then, has been a pure figment of the imagination! The statistical result of the army intelligence tests veritably "knocks the props from under" the customary 2 per cent criterion, the I. Q. of 70 as indicative of feeble-mindedness, it annihilates

the moron group, and weakens in its entirety the psychological criterion of feeble-mindedness. On the basis of army findings many may be designated "normal" mentally whose social, pedagogical, industrial, and hereditary history clearly indicates "feeble-mindedness." Whereas, heretofore, the psychological criterion has been recognized as generally the most valuable of them all, it may now become the least significant. Engaged daily and for years in the handling of social misfits, the writer will be slow to abandon established standards until the newer ones are more carefully checked both for reliability and practicability. It may be true that in the army seven to ten-year-olds functioned more or less efficiently, and that comparatively low intelligences on the army tests reported successful living at-large, yet clinical psychologists will recall few, if any cases, testing seven to ten years on the Binet who manifest such successful adaptations to the complexities of their environment. It is true, of course, that because of the nature of our work we manage to learn only about the *unsuccessful* morons. Much more satisfactory evidence must be presented, however, to convince one that the *successful* moron requires no supervision and direction, and actually guides his affairs with ordinary prudence.

(8) Because of the constant need for substitutes for the Binet test on account of previous examinations, possible coaching, etc., the writer has had more than two years' experience applying group tests, such as the Army Alpha, the Haggerty, the Illinois Examination, the National Intelligence Tests, and so on, to individual subjects. It has been amazing to find, after the pre-

scribed instructions had been read, the frequency with which additional explanations had to be given before the subject could proceed intelligently. There have been occasions where a whole test — for example, the disarranged sentences (true and false), of Army Alpha — had been entirely misunderstood. In this instance all items were underlined "false." The experimenter then explained the problem more thoroughly, the subject this time placing on record a truer representation of his degree of mental ability. Such procedure in group testing is, of course, impossible. Set directions are prescribed. If they are understood, well and good; if not, the matter cannot be helped. No questions can be entertained. It is a credit to all the group tests now in use that the directions are remarkably clear and understandable. But in every group of two hundred, a few will be found who will not quite get the idea, and when averaging results these records give a false weight to the figures.

(9) An adequate treatise dealing with the psychology of the crowd would be a valuable asset assisting in the interpretation of the army test results. As yet such an analysis is lacking. We can only speculate upon the nature of an individual's mental composition when handled merely as one in a large group. If we were to contrast individual psychological examining with group testing, it would not be unfair, perhaps, to state that the circumstance of the crowd is not conducive in all cases to maximum mental effort. There are a host of subtle factors brought into play which it has not yet been possible for scientific psychology to catch hold of. Some

instincts such as rivalry may make for extraordinary exertion and increased efficiency; on the other hand, there are other counteracting instincts which might just as likely lead to decreased efficiency. In what direction the total balance lies, whether favorable or unfavorable, cannot be dogmatically stated. It is natural to expect that under the best conditions one cannot exceed one's native ability. On the other hand, native ability may not become manifest under certain restricted conditions.

(10) The most important check upon the results of the army testings were the Stanford-Binet ages of 653 draftees from nine different camps. This group has previously been mentioned, and was designated in the army statistics as Group X. The supposition was made that this is a random sampling and is typically representative. There are a number of reasons why this sampling should be questioned. These follow:

(*a*) It is significant that with a possible maximum score of 414 in Army Alpha (weighted) no one of this group received a higher score than 360, and with a possible maximum of 212 (unweighted) no score was higher than 180. (P. 779–780.)[1]

(*b*) No member of Group X had a schooling record beyond four years of college. It is admitted that the upper range, both for mental ability and for schooling, is not here represented. (P. 779.)

(*c*) Men of foreign birth had been eliminated. (P. 780.)

(*d*) The total group (653) included a number of illiterate men. (P. 781.)

[1] These page references are to R. M. Yerkes: "Psychological Examining in the United States Army."

(*e*) Group X failed to include men who were selected as officers or for prospective officer-material. And because of the operation of the draft it was inevitable that many men, because of physical or mental handicaps, were thrown in the discard and never reached a training camp. To what extent a normal distribution is disturbed by these selective processes is a matter still to be determined.

(*f*) This group is admitted defective as a sample on p. 384. "In the first place, the group chosen in any given camp cannot be thought of as typical of the camp; and in the second place, there was the selection involved in the choice of camps."

(*g*) The attempt was not made to set down first the earmarks of a satisfactory sampling and then adhere closely to the prescription. No effort was made for a proportionate selection, covering the whole range of mental ability. It was felt, rather, that the heterogeneity of the group was a sufficient guarantee of the adequacy of the sampling. (P. 779.)

Whatever averages are obtained from Group X must be interpreted with the above facts in mind.

(11) The pragmatic test has supported the 16-year criterion. The calculation of the I. Q.'s of adults and the classifications inevitably resulting from the use of this standard have met with rather general satisfaction among clinical psychologists. The utilization of test statistics gathered in the army for the interpretation of mental capacity is well illustrated in the following report[1]

[1] From the *Survey* of Saturday, Oct. 30, 1920 (Vol. XLV., No. 5), pp. 147–148.

which is quoted in full. We may naturally expect very strange generalizations such as are here propounded:

"Prisoners vs. 'Men Generally'

"One of the most significant contributions made to the science of penology since Dr. Charles Goring's monumental report on the criminal type, seven years ago, is the recent report of the Mental Survey of Penitentiary Prisoners in Illinois. Dr. Goring's report completely repudiated the existence of a uniform physical criminal type, and established to be an illusion the doctrine which for a previous quarter of a century had done an incalculable damage to penal reform.

"This new mental survey, under the direction of Dr. Herman Adler, the criminologist, dispels the existence of a uniform mental criminal type. The basis of comparison in the Goring report was the physical characteristics of law-abiding citizens such as college students, soldiers, hospital patients, and the like; the basis of comparison in the Adler report is the result of the group mental tests applied to the United States Draft Army, comprising some 1,700,000 men of draft age from all sections and classes and racial groups in the United States. The report asserts that the result of the army test is an index of the average mental age of the people of this country.

"In the Illinois survey the conclusions drawn were on the basis of 1,650 accurate records of prisoners examined by army tests. Four and one-half per cent were in the lowest group — four per cent in the highest. In the penitentiary the inferior group was 16 per cent — in the draft army, 25 per cent. The striking fact is therefore revealed, that the penitentiary has fewer men of inferior development than the draft army, and that it has more men of superior mental development.

"The report further shows: Relatively few inferior men, relatively many superior men in the crimes of fraud

and crimes against property; the reverse in case of sex crimes; the short-termers an average group; the long-termers often either superior or inferior with a larger percentage of inferior. Those prisoners who had served previous terms had half as many inferior types, and nearly twice as many superior types as those prisoners without previous records of arrest.

"Among the reformatory prisoners the comparative results are much the same. The average age is lower and accounts for the existence of fewer superior men in the reformatory than in the draft army.

"If the analogy is correct that the American draft army is an index of the country, it is evident from this report that as a group, the prison population is not inferior to men generally."

In contrast to this analysis, in which it is demonstrated practically that penitentiary prisoners in Illinois are *superior* to men at large, are the findings of Goring,[1] of Fernald, Hayes, and Dawley,[2] of the NewYork Prison Survey Committee,[3] of Miner,[4] of Williams,[5] and of Healy.[6] The testimony of this group cannot very easily be controverted.

[1] C. Goring: "The English Convict" (abridged edition), London: His Majesty's Stationery Office, 1919, 275 pp.

[2] Mabel R. Fernald, Mary H. S. Hayes, and Almena Dawley: "A Study of Women Delinquents in New York State" (New York, Century Co., 1920, 542 pp.).

[3] Report of the Prison Survey Committee. (Albany, N. Y.: J. B. Lyon Co., 1920, 412 pp.)

[4] J. B. Miner: "Deficiency and Delinquency." (Baltimore, Warwick and York, 1918, 355 pp.)

[5] J. H. Williams: "The Intelligence of the Delinquent Boy," *Jour. of Delinquency*. Monograph 1, Jan. 1919, 198 pp.

[6] William Healy: "The Individual Delinquent" (Boston, Little, Brown & Co., 1915; 830 pp.).

The following from Goring:

"It is clear that the relationship between mental defectiveness and the committing of all types of crime, with the exception of some kinds of fraud, is an extremely intimate one. The strength of this bond transcends that of any we have hitherto been able to discover: and it is evident that defective intelligence is one of the primal sources of crime in this country." (P. 181.)

And again,

"Probably the chief source of the high degree of relationship between weak-mindedness and crime resides in the fact that the criminal thing which we call criminality, and which leads to the perpetration of many, if not of most, anti-social offenses to-day, is not inherent wickedness, but natural stupidity. At any rate, we need only study the penal records of habitual criminals to realize fully that the one characteristic of the offenses of 90 per cent of the 150,000 persons convicted to prison every year — the one characteristic, apart from their intolerableness in a well-ordered society, is the incredible stupidity of these offenses." (Pp. 182–183.)

And finally,

"Our final conclusion is that English criminals are selected by a physical condition, and a mental constitution which are independent of each other — that the one significant physical association with criminality is a generally defective physique; and that the one vital mental constitutional factor in the etiology of crime is defective intelligence." (P. 184.)

Next, Fernald, Hayes, and Dawley: (The comparison here was made direct with Group X.)

"It is evident from consideration both of the distribution and of the means that the delinquent group is inferior mentally to the army group." (P. 419.)

"The difference in mental age of 1.6 years (Group X — 13.4 years, New York delinquent women — 11.8 years) may be accepted as valid, beyond any reasonable question, since it amounts to more than ten times the standard deviation of the difference." (P. 420.)

And in summary, remembering that these statements refer only to white English-speaking groups, "the average mental capacity of the delinquent women whom we have examined is lower than that of any groups of non-delinquent adults with regard to whom we have data," and "the above statement does not imply a selection of individuals entirely from the lower end of the scale of intelligence for the delinquent group. There is, in fact, an extensive amount of overlapping of the delinquent with the non-delinquent groups." (P. 433.)

And in conclusion, "It is evident from the foregoing statement that our findings are in accord with Goring's as regards the fact of a difference and its direction. They indicate, however, a slighter degree of difference than he implies." (P. 434.)

Of course, this smaller difference is inevitable in view of the reduction of the standard.

These quotations are from the Prison Survey Committee of New York:

"A large section of the prison population consists of custodial rather than punitive cases. The prisoner who is mentally or physically unable to cope with the conditions of a free society should not be allowed to complete the cycle of commitment, release, and recommitment, and indulge between times in a criminal career as the only means within his knowledge of obtaining a livelihood. Dr. Glueck, after a study of 602 cases at Sing Sing Prison, found that 28.1 per cent showed intelligence equivalent to that of the average American child of

twelve years or under.[1] Examinations made in various prisons and reformatories, as shown by the State Prison Commission for 1918, estimated from 15 per cent to 25 per cent of the prison population as segregable on account of mental deficiency. These figures, taken in connection with the fact that the same report states that 87 per cent of the felons admitted to the prisons in 1917 had served previous terms, make obvious the fact that a large percentage of those who are released are incapable, either from certain mental defects or from lack of training in the institutions, of maintaining themselves in society. It is believed that before any great advance can be made in the training of prisoners the mental deficients must be sorted out, segregated, and dealt with as a separate unit. Elementary, advanced, and industrial education, wage incentives, self-government, and parole measures will be of little avail with this class of prisoners, whose mental limitations prevent them from taking advantage of any privilege or opportunity." (P. 104.)

One is interested to inquire to what extent this generalization applies to approximately half the American population, for the data of Group X, if regarded as representative of the country at large, indicates that only 52.7 per cent of American citizens are 13 years mentally or over! And again on p. 114 we find the remark, "From the figures obtained from various penal and correctional institutions, and from the examinations made by the Prison Survey Committee, it is believed that from 8 to 14 per cent of the prisoners, other than those in insane hospitals, are mental defectives requiring institutional care."

And now these remarks of Miner, whose regard for careful statistical analyses is a high guarantee of the soundness of his generalizations: "In the face of the fact

[1] Those of Group X under 13 years mentally amounted to 47.2%. Almost twice as many!

that mental deficiency is undoubtedly the most important single factor to be considered today in the institutional care of delinquents, one hesitates to correct even the most exaggerated impressions as to its importance." (P. 167.)

"On the basis of our summary of tested delinquents in the last chapter it seems extremely conservative to suppose that 10 per cent of the manifest and potential criminals are as deficient mentally as the lowest 1.5 per cent of the general population. Even with this assumption, we find that the chances would be 48 out of a hundred that a person of this degree of deficiency would be convicted of crime." (P. 216.)

"The most significant fact demonstrated by the correlations between juvenile delinquency and deficiency is that there is a positive relationship which is significant in amount. With the maximum estimate the correlation is nearly six times its error. This is the first time that the relationship has actually been calculated in connection with any group of juveniles." (P. 223.)

And in his conclusions, "We have summarized some of the best and most recent investigations in which a notable advance toward solving this problem has been made by means of the correlation method. This has proved to be a new and vigorous force for directing social progress. By no other method have we approached so near the solution of the cause of delinquency. It enables us to restate the problem of criminality as mainly a problem in the treatment of a hereditary criminal diathesis, in which mental deficiency is the largest factor." And then by way of a final word, "Unless the present evidence, however, is outweighed by improved data obtained in the future, the most strategic point for attacking persistent delinquency is through the relation to deficiency, with heredity holding the heights." (Pp. 244–245.)

The following are four of the conclusions of Williams:

"(2) That the general level of intelligence among delinquent, dependent, and potentially delinquent boys is decidedly lower than that of ordinary children and adults of the same ages.

"(3) That feeble-mindedness is much more common among delinquent, dependent, and potentially delinquent boys than in the population as a whole; and that approximately 30 per cent of the delinquent and dependent boys included in this investigation are definitely feeble-minded.

"(4) That low intelligence among delinquent boys is the chief contributing factor in their delinquent conduct.

"(5) That any level of intelligence lower than that of the average-normal accounts in part for delinquency, the extent to which it is responsible depending upon the degree of intelligence, which may be best expressed by the intelligence quotient." (P. 181.)

For 1000 young repeated offenders Healy reports "as beyond peradventure feeble-minded, we found about 10 per cent, but the figure will be increased as some of the younger in the lower groups fail to advance with age." (P. 140.)

And again, "The subject of mental defect is of great import in the study of delinquency and its causation. Just what percentage of delinquents are feeble-minded appears to be a matter of perennial interest, but well-founded statistics, even if obtained in particular places, may not be applicable to different situations. There can be no doubt that separate reformatory or prison populations if tested would show from 10 to 30 per cent, or even more, to be feeble-minded." (P. 447.)

"But the gist of the situation is that mental defect forms the largest single cause of delinquency to be found by correlating tendency to offend with characteristics of the offender." (P. 447.)

To return to the report in *The Survey*, in which one is given to understand that the penitentiary prisoners of Illinois are equal to men generally in mental capacity, if not slightly superior. Analyses such as these contradict all previous experience, and their cumulative force will bring about a distorted perspective, which may only confound the layman and the prison administrator in the adequate handling of our anti-social and unsocial classes.

(12) The outstanding advantage of an individual test lies in the fact that a *rapport* between examiner and examinee may be established. This is of unquestioned value in increasing a subject's efficiency and in interpreting the results of any given performance. In an individual examination failures can be explained more adequately, interest, application and distractions can be observed, *malaise* or other discomforts irritating the subject may be properly discounted. In a group test this is not possible. Your achievement within the forty-five minutes is the measure of your intelligence. No allowances can be made for any disturbing or distracting influences. To what extent the army averages have been reduced because of this inaccuracy cannot be even guessed. Combined with other reducing factors, the surprisingly low average mental age found is, to a certain extent, to be expected.

(13) Applying the "test of common sense" to the 13¼ standard, although unscientific, it possesses, nevertheless, a certain degree of validity. So-called "common sense" is generally a combination of attitudes or points of view based upon long experience, both of the individual and of the group in which an individual finds himself.

Common sense is opposed to the conception that mental *power* reaches its maturity at 13¼ years. Even now when we tell the layman that intelligence maturity is generally reached at 16, he is shockingly surprised. In reality, the whole problem of determining at what age children reach intelligence maturity is still one awaiting solution. It is realized that this objection to the 13¼ year standard is not a very weighty one, yet common sense should be considered as one among a number of the acid-tests of truth.

(14) Numerous factors of personality and temperament probably had varying influences upon the result. Such elements as ambition, conscientiousness, obedience, application, interest, were forces of variable manifestation in some manner leaving an impress upon performance. The assumption that there was an ultimate balance, excesses discounting deficiencies, is perhaps not warranted in view of the circumstance that one cannot achieve a high score on the basis of any other condition than actual mental ability.

(15) Thus far statistical sanction (the normal probability curve) has indicated no error in the 16-year criterion. If 13¼ is the average mental age of adults, shall we find the same *normal* distribution of mental ages and intelligence quotients with the 13¼ year criterion? Of course, some raise the question why we should expect a normal distribution even with the 16-year criterion.

Taking the data of Table 333 (p. 790),[1] no mental age

[1] R. M. Yerkes: "Psychological Examining in the United States Army."

of Group X exceeds 19.9 years. That is the result of the Binet examination of 653 men. For the white draft, 93,955 men, Groups I, II and III (group tests) no mental age exceeds 22.9 years. According to the evidence of the Stanford-Binet Test, intelligence of 20 or over should be found with the following relative frequencies:

MENTAL AGE	I. Q. INTERVAL	FREQUENCY IN 1000
$20\frac{1}{6}$–$21\frac{3}{5}$	126–135	16
$21\frac{5}{6}$–$23\frac{1}{6}$	136–145	2

Unless there is a great reduction in average intelligence (from 16 to 13) with increase of life age, these ratios of mental power (I. Q.) *should* be found in a correct random sampling of people at large.

Actually, the highest I. Q. using $13\frac{1}{4}$ as a base does not exceed about 147 for Group X, and for the white draft, Groups I, II and III, it does not exceed about 174. (From data of p. 790.)

(16) In order to determine the range of mental ages characteristic of adults we may be compelled to fall back upon some other measuring device than the Binet scale thus far standardized. It is inevitable, for example, that no member of Group X will test above 19.9, for the reason that a perfect performance of all the Stanford-Binet items can yield only a mental age of 19 years. A peculiar idiosyncrasy which the writer has not seen mentioned in the literature is the possibility of a slightly inferior performer obtaining some mental rating out of a possible maximum of $19\frac{1}{2}$ years, whereas one who passes all the tests in the whole series successfully

achieves a maximum of only 19. Thus with a basal age of 14, a subject is confronted with 6 tests in age 16, each having a value of 5 months, total 30 months; and 6 tests in age 18, each having a value of 6 months, total 36 months. The grand total is 66 months, or 5 years and 6 months above the basal age 14. This would yield a maximum of 19½ years. On the other hand, with a basal age of 16, a subject is confronted with only 6 tests, each having a value of 6 months, total 36 months, and the maximum is only 19 years. Until the Binet scale is more effectively extended into the upper ages, our knowledge of the nature and range of mental ability of adults will be dependent upon a certain amount of speculation.

(17) On the basis of the army statistics less than 15 per cent of the draftees examined measured over 16 years in intelligence. It is difficult to conceive in the light of the writer's ten years' testing experience that 85 per cent of adults examined have been below average and that only 15 per cent have been above average. If memory is not a deceptive servant, experience seems to point toward an average closer to 16 than to 13¼. Whether the experience of other examiners corroborates this impression cannot be determined without a proper questionnaire.

(18) The argument may be made that the army averages for intelligence are valid because the exemption of the feeble-minded counterbalances the elimination of the superior men. This is questionable for the reason that only the extremest variety of inferior ability was eliminated, whereas only a fair degree of superiority

was both recognizable as a possibility and a stimulus for officership, and in civil life for exemption because of indispensable social service. The chances are that the processes of selection which yielded Groups I, II, III and X favored the selection of those who, on the whole, were somewhat below the average of the population at large.

(19) A satisfactory standard of the average mental age of adults should include as well statistics gathered from the mental examination of women. These additional figures would serve both as a balance and as a check upon the averages obtained from the examination of men.

(20) Finally, the danger of accepting the army figures on their face value is expressed on p. 790[1] as follows: "The former group is large (Groups I, II and III) and representative, but involves an error dependent on the fact that these men were examined by alpha and beta and not by a mental age scale. The second group (Group X) suffers from the fact that it is small and cannot be demonstrated to be representative."

In conclusion, the writer wishes to express again his keen appreciation of the splendid contributions made by the psychologists during the war, regretting that conditions made it impossible for him to share in that epoch-making work. The above discussion has been motivated only by a desire to reduce the inevitable difficulties of clinical psychologists, should we hastily accept the 13¼ age standard. Caution is advised until this question

[1] R. M. Yerkes: "Psychological Examining in the United States Army."

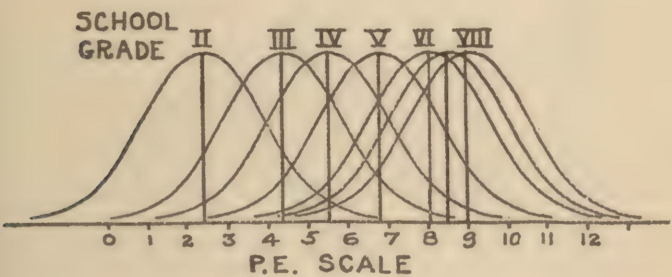
of the average mental age of adults, now that it has been so forcibly raised, is more satisfactorily settled. In the meantime, the ends of clinical and applied psychology may be more adequately served, and those engaged in the handling of our unsocial and anti-social classes may find themselves functioning more efficiently, by adhering to the 16-year standard, and by making their psychological and clinical deductions on that basis.

2. INCREASING DIVERGENCE VS. PARALLEL PROGRESS OF MENTAL DEVELOPMENT

A great deal of effort is being spent at the present time to determine the constancy of the rate of intelligence development (constancy of I. Q.). No new material will be here presented to enlighten this controversy, except to advise the utilization of a more efficient method for studying rate changes. In all of the discussions thus far, use has been made of equally-spaced cross-section paper in order to graph growth and I. Q. changes. This has resulted in greatly hindering clear analysis and, to an extent, has tended toward the development of misleading conclusions. A new method for graphing these ratios is recommended.

Psychological and educational studies seem to indicate that age and variability are positively correlated. With a given increase in age we observe a corresponding increase in variability. But when the nature of this variability has been graphed a number of inaccuracies have inevitably developed.

To illustrate with concrete examples: In his analysis of the arithmetic abilities of children, Woody[1] states, "In the construction of these scales, it has been assumed that achievement in the solution of problems in the fundamental processes is distributed according to the normal surface of frequency. Furthermore, it has been assumed that the variability of any grade from the second to the eighth is equal to that of any other." (P. 30.) Graph 45 illustrates this hypothesis.



GRAPH 45

RELATIONS OF GRADE DISTRIBUTIONS TO EACH OTHER IN ADDITION

Plotted in somewhat different fashion, Graph 46 is obtained. It will be observed that progress follows parallel lines. Variability does not increase, and presumably, children of low ability in all the grades will manifest the same difference in comparison with those of high ability, throughout. This is one point of view:

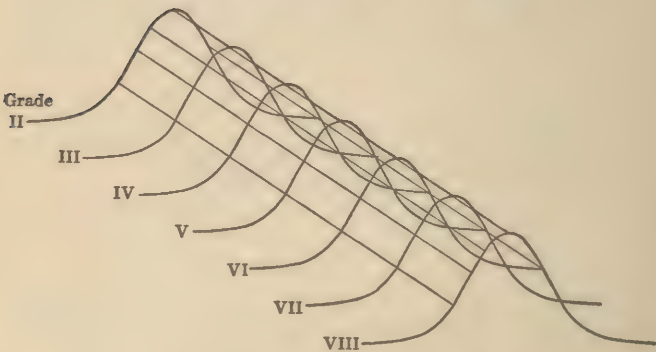
[1] C. Woody: "Measurements of Some Achievements in Arithmetic." Teachers College, Columbia University Contributions to Education, No. 80, 1916, 63 pp.

that there are mental traits which manifest parallel progress with increasing age.

On the other hand, we have the view that there is an

GRAPH 46

Parallel Progress of Mental Traits



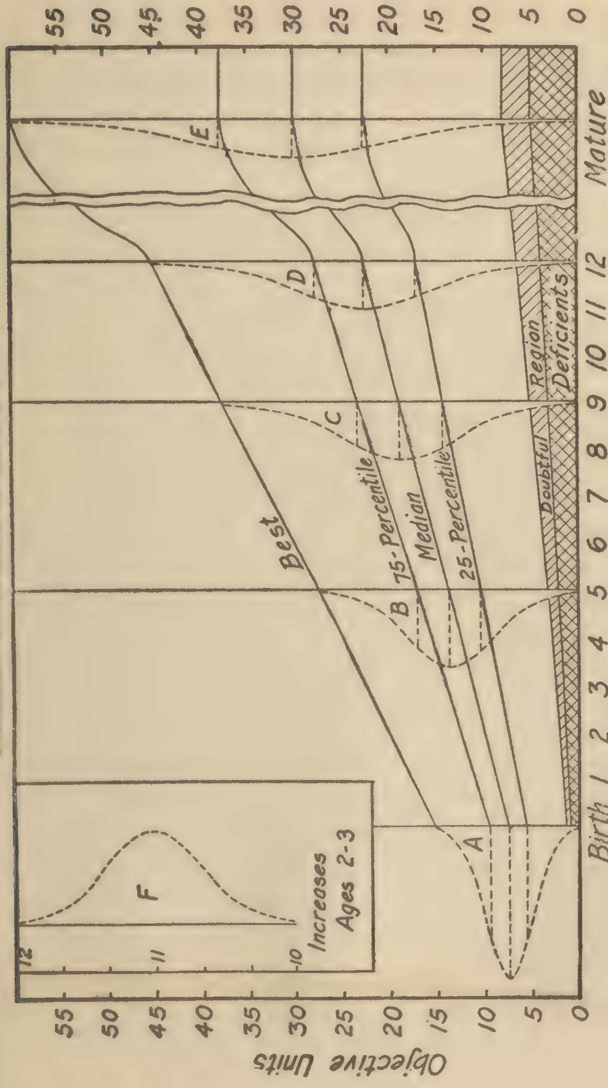
increasing divergence of mental traits with increasing age. This is graphically represented by Miner in his discussion of intelligence development.[1] (P. 253.) It will be evident from these curves that the bright and dull children deviate more and more from the average with increase of age.

To which mode of development does intelligence compare? Are there different endowments at birth presenting a normal curve of distribution, this curve being merely reduplicated at higher ages, as in Graph 46; or are there different endowments at birth presenting a

[1] J. B. Miner: "Deficiency and Delinquency."

GRAPH 47

INCREASING DIVERGENCE OF MENTAL TRAITS



Hypothetical Development Curves (Normal Distribution)

Permission Warwick & York.

normal curve of distribution, which upon increase of life age becomes more and more flat as in Graph 47? If the I. Q. is constant, then Graph 46 should adequately picture I. Q. distributions at various ages. But if for the brighter children the I. Q. increases, and for the feeble-minded it decreases, Graph 47 would more adequately represent the true state of affairs.

Studies of rate of development will not proceed toward a clear analysis of the situation until the method of graphic presentation changes from the use of equally-spaced cross-section paper to that of logarithmically-spaced cross-section paper.

Graph 49 has been derived from Fig. 10 (p. 75) of Doll's monograph,[1] only the curves for ages 3, 5, 8 and 10 being transcribed. The curves as originally presented are indicated in Graph 48. The outstanding features of difference between the two methods of presentation are so obvious that they do not require description. It will be apparent in Graph 49 that (1) I. Q. constancy at all intelligence levels, at all life ages, and at all mental ages produces a line possessing a 45 degree angle. (2) If the I. Q. increases, the angle inevitably increases. (3) If the I. Q. decreases, the angle inevitably decreases. (4) These increases and decreases appear in the graph in correct, not distorted ratios. (5) The angle at which growth proceeds is both an index of I. Q. constancy as well as a direct expression of rate of growth.

By comparing Graph 49 with Graph 48 it will be observed that on evenly-spaced cross-section paper various growth curves must be interpreted in terms of the re-

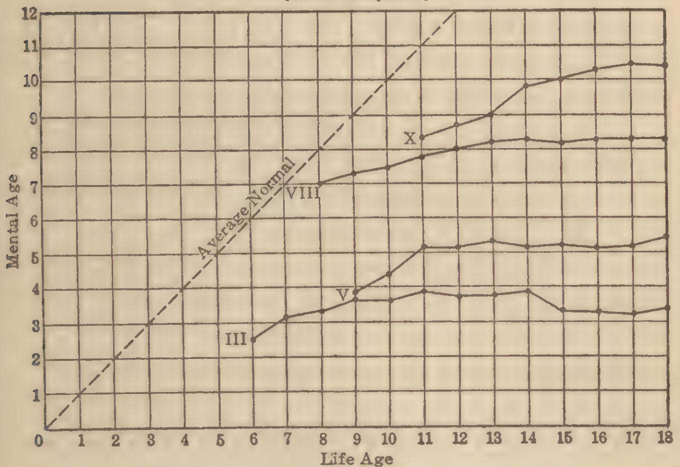
[1] E. A. Doll: "The Growth of Intelligence."

spective slopes of specific I. Q.'s, each I. Q. rate having a different slope. In Graph 49, however, all slopes, no matter what the I. Q., if indicative of constancy of development, are at a 45-degree angle. On equally-spaced charts uniformity in rate-growth would have to

GRAPH 48

Average Growth Curves of Each Mental Age (After Doll)

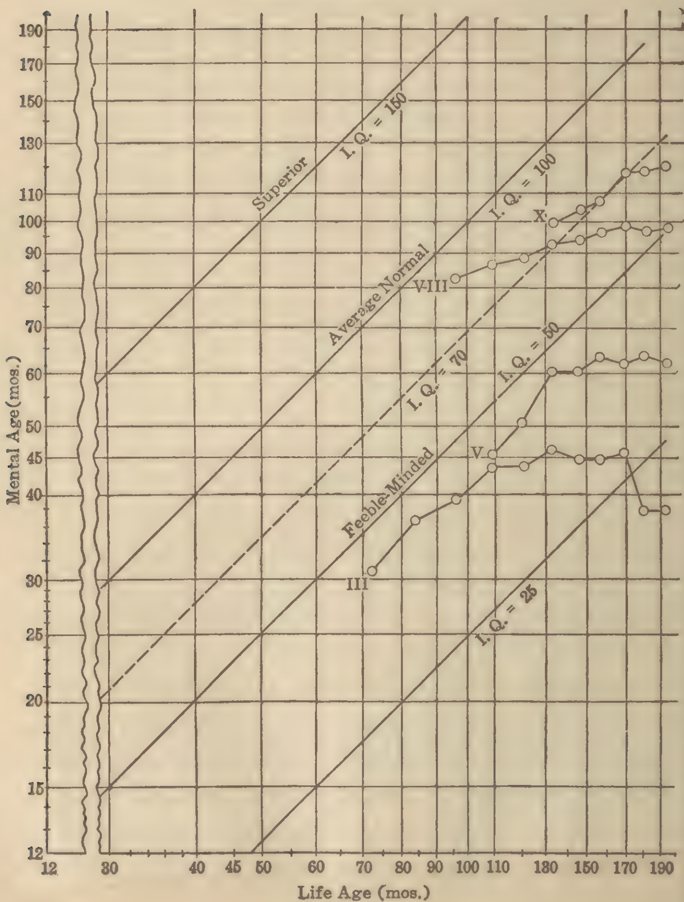
(F. M. Subjects)



be represented by different algebraic formulæ. This has an important bearing upon the interpretation of comparative curves. It appears, upon a superficial examination of the four curves of Graph 48, that there is practically little variation between them, the curves for VIII and X being almost exact duplications of the curves for III and V. Yet when VIII and X are compared with III and V in Graph 49, no such striking

GRAPH 49

Mental Age Growth Curves



similarity is evident. In Graph 48, curves III and VIII, especially between the ages of 8 to 14 are almost parallel. The actual differences are brought out more definitely in Graph 49, especially for ages 14, 15 and 16. Taking the individual curves, it is more evident in Graph 49 that age X manifests a regular reduction of I. Q. from age 11 to 16; that age VIII also manifests a constant regular reduction of I. Q. somewhat greater than age X; that age V shows a slight increase in rate of growth up to 132 months (age 11), then there is a rapid regression, the ratio of which remains the same up to age 16; and that age III manifests only a slight decrease in rate of growth up to 108 months (age 9), then follows a sudden constant rate of regression up to 168 months (age 14), then the reduction takes a rapid increase up to age 16.

Reference to Fig. 12 (p. 107) of Doll's monograph will emphasize again the need for this suggested method for graphing growth curves. In this figure are presented the individual growth curves showing early arrest with consequent I. Q. decreases. All these curves (Cases 9, 17, 16, 48, 64, 98, 134, 164) are practically parallel, and it was necessary to indicate under each curve the different I. Q. values at each life age. The differences of I. Q. at the two ends of each curve are here presented:

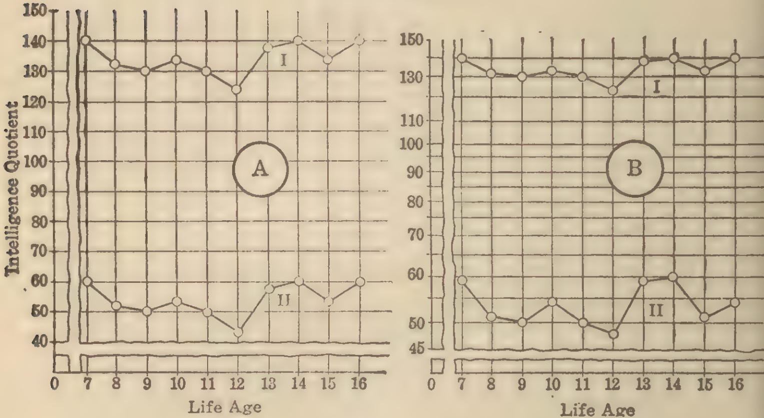
Case No.	I. Q. Range	Difference
9	25–17	12
17	38–25	13
16	51–38	13
48	47–37	10
64	62–46	16
98	70–46	24
134	91–55	36
164	93–75	18

On logarithmic cross-section paper these differences would have been more apparent.

Graph 50 illustrates, in another fashion, the advantages of the ratio graph over the customary method of plotting rate progress. Two curves are here presented, labeled I and II. The two curves run exactly parallel

GRAPH 50

I. Q. Fluctuations



in (*A*), yet in (*B*), where rate is more adequately represented, no such parallel progress is observed. It is a matter of common knowledge that at lower ages, because of the exaggerated significance of small deviations, there is bound to be greater inconstancy. A deviation of 6 months at age 6 will produce a marked change of I. Q. At age 16, however, a fluctuation of that size is relatively insignificant. To what extent this condition accounts for some of the spurious inconstancies at lower

ages is yet to be determined. It is probable that cross-section paper reversing the divisions, so that at higher ages fluctuations of a given magnitude are granted greater weight, and at lower ages, less weight, may equalize the inevitable discrepancies.

One of the important tasks confronting experimental psychologists who are investigating the problem of intelligence development is to determine the nature of its progress. Do endowments unfold or develop at a uniform rate, or are there increments or decrements, regular or irregular? To answer this question, perhaps a variety of intelligence tests, other than the Binet, may have to be utilized. And the statistical interpretation of the problem will be greatly facilitated by the use of the ratio chart.[1]

3. A PERFORMANCE INTELLIGENCE SCALE

The work of the last few years has contributed considerable material on performance intelligence tests. The block designs, it is hoped, will be found to be a practicable addition thereto. Although this test may be utilized as a measure of intelligence by itself, its greatest value, however, may be realized in combination with other performance tests which are capable of incorporation in such a scale.

The time is ripe for the development of a performance intelligence scale comparable in construction to the Binet. An especially rich field for data is the monograph

[1] Irving Fisher: "The 'Ratio' Chart." Quarterly Publications American Statistical Association, 1917, pp. 577–601.

dealing with the results of intelligence tests in the army.[1]

It was not regarded within the province of this piece of research to develop such a scale, the purpose being rather to standardize a new set of tests, discussing therewith numerous allied questions and problems.

[1] R. M. Yerkes: "Psychological Examining in the Army."

APPENDIX I

Grouped Distributions of Ability

OUTLINE

A. Three Groups

- *a.* $6\,\sigma$
- *b.* $5\,\sigma$
- *c.* 12 Q

B. Five Groups

- *a.* $6\,\sigma$
- *b.* $5\,\sigma$
- *c.* 12 Q

C. Seven Groups

- *a.* $6\,\sigma$
- *b.* $5\,\sigma$
- *c.* 12 Q

D. Nine Groups

- *a.* $6\,\sigma$
- *b.* $5\,\sigma$
- *c.* 12 Q

E. Practical Applications

- *a.* School Marks
- *b.* Economic Demand
- *c.* Testing Psychological Tests

GROUPED DISTRIBUTIONS OF ABILITY

It is the purpose of this section to assist those attempting to determine how accurately a test or scheme of tests approximates normal probability in its distribution of results, when the effort is made to classify the individuals measured into three, five, seven, or nine groups. The discussion of groupings to determine which leads to most satisfactory results will be deferred for later treatment.

A. *Three Groups*

Let us assume that we wish to classify individual differences in mental ability or in any trait or characteristic into three groups. What proportion of 100 per cent would we expect to find in each group?[1]

Let us consider three possibilities, where the limits of the distribution-range are

(*a*) $\pm 3\,\sigma$
(*b*) $\pm 2.5\,\sigma$
(*c*) ± 6 Q.

(*a*) $\pm 3\,\sigma$. Range is $6\,\sigma$.

Since we are dealing with three groups, each will have an extent of $6/3\,\sigma$, or $2\,\sigma$. Consequently the limits for the three groups will be

First Group,	.00 to 2.00
Second Group,	2.01 to 4.00
Third Group,	4.01 to 6.00

[1] The final figures in this article are obtained from the formula of the normal probability curve,

$$y = \frac{n}{\sigma\sqrt{2\pi}} \cdot \frac{1}{e^{\frac{x^2}{2\sigma^2}}}$$

From a table[1] giving the values of x/σ we find that the

first group includes 15.73% of cases,
first and second groups include 84.27% of cases.

Therefore:

1st Group includes 15.73% of cases
2nd Group includes 68.54% of cases
3rd Group includes 15.73% of cases

(*b*) $\pm 2.5\,\sigma$ Range is $5\,\sigma$

Since we are dealing with three groups, each will have an extent of $5/3\,\sigma$, or $1.667\,\sigma$. Consequently the limits for the three groups will be

First Group,	.000 to 1.667
Second Group,	1.668 to 3.333
Third Group,	3.334 to 5.000

From the x/σ table[2] we find that the

first group includes 19.63% of cases,
first and second groups include 80.37% of cases.

Therefore:

1st Group includes 19.63% of cases
2nd Group includes 60.74% of cases
3rd Group includes 19.63% of cases

[1] H. O. Rugg, "Statistical Methods Applied to Education," 1917, Table VI, pp. 396–400.
[2] Table V, pp. 392–395, in Rugg.

(*c*) ± 6 Q. Range is 12 Q.

Since we are dealing with three groups, each will have an extent of 12/3 Q, or 4 Q. Consequently the limits for the three groups will be

First Group, 6 Q — 2 Q
Second Group, 2 (2 Q)
Third Group, 6 Q — 2 Q

From the x/P. E. table (Table IV, p. 391 in Rugg, interpolated) the percentage included within each group will be

First Group, 50.00 — 41.13%
Second Group, 2 (41.13)%
Third Group, 50.00 — 41.13%

Therefore:

1st Group includes 8.87% of cases
2nd Group includes 82.26% of cases
3rd Group includes 8.87% of cases

B. *Five Groups*

Let us assume that we wish to classify individual differences in a given trait into five groups. What proportion of 100 per cent would we expect to find in each group? Let us again consider three possibilities where the limits of the score-range are

(*a*) $\pm 3\ \sigma$
(*b*) $\pm 2.5\ \sigma$
(*c*) ± 6 Q

(*a*) $\pm 3\ \sigma$ Range is $6\ \sigma$

Since we are dealing with five groups, each will have an extent of $6/5\ \sigma$, or $1.2\ \sigma$. The limits for the five groups will be

First Group,	.00 to 1.20
Second Group,	1.21 to 2.40
Third Group,	2.41 to 3.60
Fourth Group,	3.61 to 4.80
Fifth Group,	4.81 to 6.00

From our tables we find that the

1st group includes 3.45% of cases
1st and 2nd groups include 27.29% of cases
1st, 2nd, and 3rd groups include 72.71% of cases
1st, 2nd, 3rd, and 4th groups include 96.55% of cases

Therefore:

1st Group includes 3.45% of cases
2nd Group includes 23.84% of cases
3rd Group includes 45.42% of cases
4th Group includes 23.84% of cases
5th Group includes 3.45% of cases

(*b*) $\pm 2.5\ \sigma$ Range is $5\ \sigma$

Extent of each group is $5/5\ \sigma$, or $1\ \sigma$. The limits for the five groups will be

First Group,	.00 to 1.00
Second Group,	1.01 to 2.00
Third Group,	2.01 to 3.00
Fourth Group,	3.01 to 4.00
Fifth Group	4.01 to 5.00

From our tables we find that the

1st group includes 6.06% of cases
1st and 2nd groups include 30.23% of cases
1st, 2nd, and 3rd groups include 69.77% of cases
1st, 2nd, 3rd, and 4th groups include 93.94% of cases

Therefore:

1st Group includes 6.96% of cases
2nd Group includes 24.17% of cases
3rd Group includes 39.54% of cases
4th Group includes 24.17% of cases
5th Group includes 6.06% of cases

(*c*) ± 6 Q Range is 12 Q

Extent of each group is 12/5 Q, or 2.40 Q. The limits for the five groups will be

First Group,	6.00 Q to 3.60 Q
Second Group,	3.60 Q to 1.20 Q
Third Group,	2.00 to 1.20 Q
Fourth Group,	3.60 Q to 1.20 Q
Fifth Group,	6.00 Q to 3.60 Q

From the tables we find that the percentage included within each group will be

First Group,	50.00 to 49.24%
Second Group	49.24 to 29.08%
Third Group,	2.00 to 29.08%
Fourth Group,	49.24 to 29.08%
Fifth Group,	50.00 to 49.24%

Therefore:

1st Group includes 0.76% of cases
2nd Group includes 20.16% of cases
3rd Group includes 58.16% of cases
4th Group includes 20.16% of cases
5th Group includes 0.76% of cases

C. *Seven Groups*

(*a*) $\pm 3\,\sigma$ Range is $6\,\sigma$

Extent of each group is $6/7\,\sigma$, or $.857\,\sigma$. The limits for the seven groups are:

First Group,	.000 to .857
Second Group,	.858 to 1.714
Third Group,	1.715 to 2.571
Fourth Group,	2.572 to 3.429
Fifth Group,	3.430 to 4.286
Sixth Group,	4.287 to 5.143
Seventh Group,	5.144 to 6.000

From the x/σ table we find that the

1st group includes 1.47% of cases
1st and 2nd groups include 9.78% of cases
1st, 2nd, and 3rd groups include 33.26% of cases
1st, 2nd, 3rd, and 4th groups include 66.74% of cases

Therefore:

1st Group includes 1.47% of cases
2nd Group includes 8.31% of cases
3rd Group includes 23.48% of cases
4th Group includes 33.48% of cases

5th Group includes 23.48% of cases
6th Group includes 8.31% of cases
7th Group includes 1.47% of cases

(*b*) $\pm 2.5\,\sigma$ Range is $5\,\sigma$

The extent of each group is $5/7\,\sigma$, or $.714\,\sigma$. The limits for the seven groups are

First Group,	.000 to .714
Second Group,	.715 to 1.429
Third Group,	1.430 to 2.143
Fourth Group,	2.144 to 2.857
Fifth Group,	2.858 to 3.572
Sixth Group,	3.573 to 4.286
Seventh Group,	4.287 to 5.000

From the x/σ table we find that the

1st group includes 3.08% of cases
1st and 2nd groups include 13.59% of cases
1st, 2nd, and 3rd groups include 35.43% of cases
1st, 2nd, 3rd, and 4th groups include 64.57% of cases

Therefore:

1st Group includes 3.08% of cases
2nd Group includes 10.51% of cases
3rd Group includes 21.84% of cases
4th Group includes 29.14% of cases
5th Group includes 21.84% of cases
6th Group includes 10.51% of cases
7th Group includes 3.08% of cases

(*c*) ± 6 Q Range is 12 Q

The extent of each group is 12/7 Q, or 1.71 Q. The limits for the seven groups will be

First Group,	6.00 Q – 4.285 Q
Second Group,	4.285 Q – 2.571 Q
Third Group	2.571 Q – .857 Q
Fourth Group,	2 (.857 Q)
Fifth Group,	2.571 Q – .857 Q
Sixth Group,	4.285 Q – 2.571 Q
Seventh Group,	6.00 Q – 4.285 Q

From the tables we find that the percentage included within each group will be

First Group,	50.00 to 49.81%
Second Group,	49.81 to 45.85%
Third Group,	45.85 to 21.84%
Fourth Group,	2 (21.84%)

Therefore:

1st Group includes	*.19% of cases*
2nd Group includes	*3.96% of cases*
3rd Group includes	*24.01% of cases*
4th Group includes	*43.68% of cases*
5th Group includes	*24.01% of cases*
6th Group includes	*3.96% of cases*
7th Group includes	*.19% of cases*

It is recommended that wherever possible larger group differentations be utilized in educational and psychological investigations. This practice not alone yields more intelligible material but is of superior service in correlational analyses.

D. *Nine Groups*

(*a*) $\pm 3\,\sigma$ Range is $6\,\sigma$

Extent of each group is $6/9\,\sigma$, or $.667\,\sigma$. The limits for the nine groups are

First Group,	.000 to .667
Second Group,	.668 to 1.333
Third Group,	1.334 to 2.000
Fourth Group,	2.001 to 2.667
Fifth Group,	2.668 to 3.333
Sixth Group,	3.334 to 4.000
Seventh Group,	4.001 to 4.667
Eighth Group,	4.668 to 5.333
Ninth Group,	5.334 to 6.000

From the x/σ table we find that the

1st group includes .84% of cases
1st and 2nd groups include 4.64% of cases
1st, 2nd, and 3rd groups include 15.73% of cases
1st, 2nd, 3rd, and 4th groups include 36.82%
1st, 2nd, 3rd, 4th, and 5th groups include 63.18%

Therefore:

1st Group includes .84% of cases
2nd Group includes 3.80% of cases
3rd Group includes 11.09% of cases
4th Group includes 21.09% of cases
5th Group includes 26.36% of cases
6th Group includes 21.09% of cases
7th Group includes 11.09% of cases
8th Group includes 3.80% of cases
9th Group includes .84% of cases

(*b*) $\pm 2.5\,\sigma$ Range is $5\,\sigma$

The extent of each group is $5/9\,\sigma$, or $.556\,\sigma$. The limits for the nine groups are

First Group,	.000 to .556
Second Group,	.557 to 1.111
Third Group,	1.112 to 1.667
Fourth Group,	1.668 to 2.222
Fifth Group,	2.223 to 2.778
Sixth Group,	2.779 to 3.333
Seventh Group,	3.334 to 3.889
Eighth Group,	3.890 to 4.444
Ninth Group,	4.445 to 5.000

From the x/σ table we find that the

1st group includes 1.98% of cases
1st and 2nd groups include 7.63% of cases
1st, 2nd, and 3rd groups include 19.63% of cases
1st, 2nd, 3rd, and 4th groups include 38.43% of cases
1st, 2nd, 3rd, 4th, and 5th groups include 61.57% of cases

Therefore:

1st Group includes 1.98% of cases
2nd Group includes 5.65% of cases
3rd Group includes 12.00% of cases
4th Group includes 18.80% of cases
5th Group includes 23.14% of cases
6th Group includes 18.80% of cases
7th Group includes 12.00% of cases
8th Group includes 5.65% of cases
9th Group includes 1.98% of cases

(*c*) ± 6 Q Range is 12 Q

The extent of each group is 12/9 Q, or 1.33 Q. The limits for the nine groups will be

First Group,	6.00 Q – 4.667 Q
Second Group,	4.667 Q – 3.333 Q
Third Group,	3.333 Q – 2.000 Q
Fourth Group,	2.000 Q – .667 Q
Fifth Group,	2 (.667 Q)
Sixth Group,	2.000 Q – .667 Q
Seventh Group,	3.333 Q – 2.000 Q
Eighth Group,	4.667 Q – 3.333 Q
Ninth Group,	6.00 Q – 4.667 Q

From the tables we find that the percentage included within each group will be

First Group,	50.00 to 49.91%
Second Group,	49.91 to 48.78%
Third Group,	48.78 to 41.13%
Fourth Group,	41.13 to 17.36%
Fifth Group,	2 (17.36%)

Therefore:

1st Group includes	*.09% of cases*
2nd Group includes	*1.13% of cases*
3rd Group includes	*7.65% of cases*
4th Group includes	*23.77% of cases*
5th Group includes	*34.72% of cases*
6th Group includes	*23.77% of cases*
7th Group includes	*7.65% of cases*
8th Group includes	*1.13% of cases*
9th Group includes	*.09% of cases*

In the following tables are given the various percentages for three, five, seven, and nine groups, where the ranges are 6, 5, and 12 Q respectively.

TABLE LXXXV
RANGE 6 σ

Group	Below			Middle			Above			Total
	4	3	2	1	0	1	2	3	4	
3				15.73	68.54	15.73				100%
5			3.45	23.84	45.42	23.84	3.45			100%
7		1.47	8.31	23.48	33.48	23.48	8.31	1.47		100%
9	.84	3.80	11.09	21.09	26.36	21.09	11.09	3.80	.84	100%

TABLE LXXXVI
RANGE 5 σ

Group	Below			Middle			Above			Total
	4	3	2	1	0	1	2	3	4	
3				19.63	60.74	19.63				100%
5			6.06	24.17	39.54	24.17	6.06			100%
7		3.08	10.51	21.84	29.14	21.84	10.51	3.08		100%
9	1.98	5.65	12.00	18.80	23.14	18.80	12.00	5.65	1.98	100%

TABLE LXXXVII
RANGE 12 Q

Group	Below			Middle			Above			Total
	4	3	2	1	0	1	2	3	4	
3				8.87	82.26	8.87				100%
5			0.76	20.16	58.16	20.16	0.76			100%
7		0.19	3.96	24.01	43.68	24.01	3.96	0.19		100%
9	0.09	1.13	7.65	23.77	34.72	23.77	7.65	1.13	.09	100%

The values in the body of the tables are percentages.

E. *Practical Applications*

1. School Marks: In assigning grades to pupils teachers often utilize a five-group classification, the range being assumed 5 σ. Thus Starch in his "Educational Measurements" states: "On the basis of all these experimental results it would seem that the most satisfactory marking system would be a scale of five steps:

A, or Excellent which should be assigned to approximately 7% of the pupils.
B, or Superior, which should be assigned to approximately 24% of the pupils.
C, or Average, which should be assigned to approximately 38% of the pupils.
D, or Inferior, which should be assigned to approximately 24% of the pupils.
E, or Unsatisfactory, which should be assigned to approximately 7% of the pupils." (P. 15.)

And Rugg in his "Statistical Methods Applied to Education" (p. 219) gives the same figures.

The chief advantages of this system of marking are, first, that the theoretical values for each group are based upon correct statistical theory; second, the groups — superior, average, inferior, etc. — cover equal ranges of ability; and third, the personal equation of the examiner is greatly reduced, if not entirely eliminated.

When the score-range or mark-range is 6 σ, or 12 Q, and should more than a five-group classification be desired, the tables in this article may be found of value.

2. Satisfying Economic Demand: In view of the fact that physical and mental traits yield distributions

closely resembling the normal probability curve, it is logical to assume that the satisfaction of needs to which these traits are allied would be compelled to conform more or less closely to this probability. For example, head circumference, length of foot, length of torso, length of hand, etc., are traits which yield Gaussian or bell-shaped distributions. Therefore, the hatter, the shoe manufacturer, the shirt manufacturer, the glover, and so on, must take account of grouped variability. If the hatter is going to manufacture nine sizes of hats and if head circumference is found distributed over a range of 6 σ, then having chosen equal units of difference between hat sizes, the per cent of each size must be as follows: (from smallest to largest size), 1, 4, 11, 21, 26, 21, 11, 4, 1.

3. Testing Tests: In standardizing tests or scales, it is desirable to determine how closely one's results conform to the normal frequency. Where one's score-range is 6 σ, or 5 σ, or 12 Q, the tables in this article readily lend themselves for such comparisons. Should the score-range be different, then a new table of values has to be derived, utilizing the same method as we have followed above. Thus, in Terman's "The Stanford Revision and Extension of the Binet-Simon Scale for Measuring Intelligence" we note a comparison of nine groups of obtained intelligence quotients with the percentage one would theoretically expect to find. The range of intelligence quotients is from 56 to 145, or about 8 Q. The obtained percentages for the nine groups are: .33, 2.3, 8.6, 20.1, 33.9, 23.1, 9., 2.3, .55 (p. 42). The theoretical values are: .16, 1.6, 8.5, 23.42, 32.64, 23.42, 8.5, 1.6,

.16. The correspondence is remarkably close and is a strong argument in favor of the validity of the intelligence quotient, as well as of the Stanford Revision of the Binet scale.

With the application of statistical methods to psychological, educational, and industrial problems becoming more common, we may expect a wider and wider use of grouped variability.

APPENDIX II

New Correlation Method

In his "Mental and Social Measurements" (2nd edition) (pp. 172–177) Thorndike points out that the formula for r stated as follows,

$$r = \frac{\Sigma(x \cdot y)}{\sqrt{\Sigma x^2} \cdot \sqrt{\Sigma y^2}}, \tag{1}$$

is a simpler form of the Pearson coefficient of correlation known as

$$r = \frac{\Sigma(x \cdot y)}{n\sigma_A\sigma_B}. \tag{2}$$

In Table 42, facing p. 176, Thorndike presents the method for determining r by formula (1).

The writer has had frequent occasion to observe that because of the elimination of the frequency interval which contains the median, together with the number of cases therein found, the normal balance of the correlation is seriously disturbed. Thus, correlation tables are met with in which the frequency is skewed and the median barely falls into the interval containing it. In spite of this condition, the entire interval is disregarded in the calculations for r. Some very weird coefficients inevitably result, differing considerably from r as would have been determined bv means of formula (2).

It was to correct this occasional inaccuracy that the following method has been used throughout this monograph.

Table LXXXVIII presents the procedure commonly practiced. Table LXXXIX illustrates the suggested method of determining the value of *r*. Scientific accuracy is increased without increase of labor.

It will be observed that the number of cases is 57, the median case being the 29th. For ability in A, the median just barely falls into the fourth frequency interval. For ability in B, the median falls well into the fourth frequency interval. Table LXXXIX takes this difference into account in the following manner: For the x variable the median falls upon the first case in 17. If this group is divided into four sections each section will contain 4¼ cases. The general rule is, when the median case fall in the first fourth section divide your table into two segments, the dividing line being drawn to the left of the frequency interval; when the median case falls within the second or third fourths (the middle half), then *only*, should the whole frequency be eliminated. In that event lines are drawn both to the right and left of the interval. If, however, the median falls in the last fourth, the table is again divided into only two segments, the dividing line being drawn to the right of the frequency interval. This procedure is illustrated in Table LXXXIX, the median for ability in A falling in the first quarter of 17, and the median for ability in B falling in the middle half of the frequency 12.

The mathematical proof has been supplied by Dr. J. E. Coover and follows the tables.

TABLE LXXXVIII
CORRELATION METHOD COMMONLY USED

Ability in A

Ability in B	1	2	3	4	5	6	7	TOTAL	
6 . . .		1	2	2	1			6	3
7 . . .	5	1	1	1				8	2
8 . . .		3	5	2				10	1
9 . . .		2	6	2	2			12	0
10 . . .		1	1	5	2			9	1
11 . . .				4	3	1		8	2
12 . . .				1		2	1	4	3
Total .	5	8	15	17	8	3	1	57	↑
$x \rightarrow$	3	2	1	0	1	2	3		y

Σx^2	Σy^2	$+\Sigma xy$		$-\Sigma xy$	
$23 \cdot 1^2 = 23$	$19 \cdot 1^2 = 19$	30	2	2	3
$11 \cdot 2^2 = 44$	$16 \cdot 2^2 = 64$	6	6	1	
$6 \cdot 3^2 = 54$	$10 \cdot 3^2 = 90$	4	4	Total = 6	
Total = 121	Total = 173	6	12		
		6	9	$\Sigma xy = 86$	
		2			
		5			
		Total = 92			

$$r = \frac{86}{\sqrt{121} \cdot \sqrt{173}} = \frac{86}{11 \times 13.2}$$

$$= \frac{86}{145.2} = +.59$$

TABLE LXXXIX

CORRELATION METHOD USED IN THIS MONOGRAPH

Ability in A

Ability in B	1	2	3	4	5	6	7	Total	
6		1	2	2	1			6	3
7	5	1	1	1				8	2
8		3	5	2				10	1
9		2	6	2	2			12	0
10		1	1	5	2			9	1
11				4	3	1		8	2
12				1		2	1	4	3
Total .	5	8	15	17	8	3	1	57	↑
$x \rightarrow$	5	3	1	1	3	5	7		y

Σx^2	Σy^2	$+\Sigma xy$		$-\Sigma xy$	
$32 \cdot 1^2 = 32$	$19 \cdot 1^2 = 19$	50	5	3	6
$16 \cdot 3^2 = 144$	$16 \cdot 2^2 = 64$	9	8	1	2
$8 \cdot 5^2 = 200$	$10 \cdot 3^2 = 90$	6	3		2
$1 \cdot 7^2 = 49$	Total = 173	9	6		9
Total = 425		6	18	Total = 23	
		2	10		
		5	30		
			21		
		Total = 188		$\Sigma xy = 165$	

$$r = \frac{165}{\sqrt{425} \cdot \sqrt{173}} = \frac{165}{20.6 \cdot 13.2} = \frac{165}{271.9}$$
$$= +.61$$

Using Pearson's longer formula (Formula 2)

$$r = \frac{82.45}{57 \times 1.36 \times 1.74} = \frac{82.45}{134.88}$$
$$= +.61$$

NOTE: Utilizing the method suggested by Thorndike (Table 42, in reference above-mentioned), r for Table LVI, p. 145, is $+ .73$; by this method r is reduced to $+ .71$

The Pearson-Bravais formula for the coefficient of correlation

$$r = \frac{\Sigma\, xy}{n\, \sigma_x \sigma_y}$$

may be converted into the formula

$$r = \frac{\Sigma\, xy}{\sqrt{\Sigma\, x^2 \times \Sigma\, y^2}}$$

in both of which,

x = the deviations from the mean of the x's,
and y = the deviations from the mean of the y's.

When the deviations are calculated from an arbitrary origin in order to keep them integral and facilitate calculation, it is customary to make corrections as shown in the formula

$$r = \frac{\frac{\Sigma\, x'y'}{n} - \nu_{1\,x}\nu_{1\,y}}{\sigma_x{}'\sigma_y{}'}$$

in which x' and y' are deviations from an arbitrary origin and σ' is the Standard Deviation in terms of the class units; $\nu_{1\,x} = \frac{\Sigma\, x'}{n}$ = the correction for the mean.

When the number of cases is small, and the correction is deemed negligible, the following formula has frequently been used:

$$r = \frac{\Sigma\, x'y'}{\sqrt{\Sigma\, x'^2 \times \Sigma\, y'^2}}$$

and the cases falling in the classes from which x' and y' have their origin are disregarded and have no influence on the calculation of the coefficient. If the means of x and y chance to fall midway in the classes chosen for the origin of their deviations, no correction would be necessary, and the coefficient would be the same as when calculated by the longer complete formula. For this reason the median class is usually chosen for the origin of deviations. If, however, either or both means chance to fall near the class-limit, the error may be appreciable.

Hence, the same formula has been used in such a way as to place the origin of deviations at either the mid-point of the median class or one of its class-limits, according to which lies nearest to the median value in the distribution. This modification decreases the possible difference between the mean and the origin of deviations from ½ to ¼ of a class-unit, and when the origin of deviations falls nearest a class-limit increases the proportion of the cases in the table that are utilized in the calculation of the coefficient.

When the origin of the deviations is placed at a class-limit, the class deviations run as follows: .5, 1.5, 2.5, etc.; decimals are eliminated by multiplying by 2; as, 1, 3, 5, 7, etc. Correction for this change in the class-deviations is automatic in the formula:

(1) Suppose origins for x' and y' both fall at class-limits, and a double-prime designates the class-deviations that are multiplied by 2.

Then

$$\Sigma\, x'y' = \frac{\Sigma\, x''y''}{2 \times 2},\ \Sigma\, x'^2 = \frac{\Sigma\, x''^2}{(2)^2},\ \Sigma\, y'^2 = \frac{\Sigma\, y''^2}{(2)^2},$$

and
$$\frac{\dfrac{\Sigma\, x''y''}{4}}{\sqrt{\dfrac{\Sigma\, x''^2}{4} \times \dfrac{\Sigma\, y''^2}{4}}} = \frac{\Sigma\, x'y'}{\sqrt{\Sigma\, x'^2 \times \Sigma\, y'^2}} = r.$$

(2) Suppose only the origin of x' falls on a class-limit:

Then
$$\Sigma\, x'y' = \frac{\Sigma\, x''y'}{2}, \quad \Sigma\, x'^2 = \frac{\Sigma\, x''^2}{(2)^2}$$

and
$$\frac{\dfrac{\Sigma\, x''y'}{2}}{\sqrt{\dfrac{\Sigma\, x''^2}{4} \times \Sigma\, y'^2}} = \frac{\Sigma\, x'y'}{\sqrt{\Sigma\, x'^2 \cdot \Sigma\, y'^2}} = r.$$

The amount of error involved in the use of the formula may be tested without recalculating by the following correction:

(*a*) For the numerator:

$$-\frac{\Sigma\, x' \times \Sigma\, y'}{n}.$$

(*b*) For the denominator:

$$\sqrt{+\left(\frac{(\Sigma\, x')^2}{n} \times \frac{(\Sigma\, y')^2}{n}\right) - \left(\Sigma x'^2 \times \frac{(\Sigma\, y')^2}{n} + \Sigma y'^2 \times \frac{(\Sigma\, x')^2}{n}\right)}.$$

APPENDIX III

Elaboration of Sheppard's Table

In order to add to the convenience of those utilizing tables of values of the normal probability integral, Sheppard's Table for Q (or P. E.), presented as Table 45 (p. 200) in Thorndike's "Mental and Social Measurements" (2d edition), and as Table XLVII (p. 116) in Buckingham's "Spelling Ability," has been elaborated, giving the straight-line interpolations to one-hundredths P. E. in Table XC following:

TABLE XC

PER CENTS OF THE NORMAL FREQUENCY SURFACE CORRESPONDING TO GIVEN VALUES OF P. E.

(The Median is the Point of Origin of P. E. Values)

$\frac{x}{\text{P. E.}}$	Per Cent of Cases									
	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.00	00.00	0.27	0.54	0.81	1.08	1.35	1.62	1.89	2.15	2.42
.10	2.69	2.96	3.23	3.49	3.76	4.03	4.30	4.57	4.83	5.09
.20	5.36	5.63	5.90	6.16	6.43	6.70	6.96	7.23	7.49	7.76
.30	8.02	8.28	8.54	8.81	9.07	9.33	9.59	9.85	10.11	10.37
.40	10.63	10.89	11.15	11.41	11.67	11.93	12.19	12.44	12.70	12.95
.50	13.21	13.46	13.71	13.97	14.22	14.47	14.72	14.97	15.21	15.46
.60	15.71	15.96	16.21	16.45	16.70	16.95	17.19	17.43	17.68	17.92
.70	18.16	18.40	18.64	18.87	19.11	19.35	19.59	19.82	20.06	20.29
.80	20.53	20.76	20.99	21.22	21.45	21.68	21.91	22.13	22.36	22.58
.90	22.81	23.03	23.25	23.48	23.70	23.92	24.14	24.35	24.57	24.78
1.00	25.00	25.21	25.42	25.64	25.85	26.06	26.27	26.47	26.68	26.88
1.10	27.09	27.29	27.49	27.70	27.90	28.10	28.30	28.49	28.69	28.88
1.20	29.08	29.27	29.46	29.66	29.85	30.04	30.23	30.41	30.60	30.78
1.30	30.97	31.15	31.33	31.52	31.70	31.88	32.05	32.23	32.40	32.58
1.40	32.75	32.92	33.09	33.26	33.43	33.60	33.76	33.92	34.09	34.25
1.50	34.41	34.57	34.73	34.89	35.05	35.21	35.36	35.51	35.67	35.82
1.60	35.97	36.12	36.27	36.41	36.56	36.71	36.85	36.99	37.14	37.28
1.70	37.42	37.56	37.70	37.83	37.97	38.11	38.24	38.37	38.50	38.63
1.80	38.76	38.89	39.01	39.14	39.26	39.39	39.51	39.63	39.76	39.88
1.90	40.00	40.11	40.23	40.34	40.46	40.57	40.68	40.79	40.91	41.02
2.00	41.13	41.24	41.34	41.45	41.55	41.66	41.76	41.86	41.97	42.07

TABLE XC — *Continued*

$\frac{x}{P.E.}$	Per Cent of Cases									
	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
2.10	42.17	42.27	42.36	42.46	42.55	42.65	42.74	42.83	42.93	43.02
2.20	43.11	43.20	43.28	43.37	43.45	43.54	43.62	43.71	43.79	43.88
2.30	43.96	44.04	44.12	44.19	44.27	44.35	44.42	44.50	44.57	44.65
2.40	44.72	44.79	44.86	44.94	45.01	45.08	45.15	45.21	45.28	45.34
2.50	45.41	45.47	45.54	45.60	45.67	45.73	45.79	45.85	45.90	45.96
2.60	46.02	46.08	46.14	46.19	46.25	46.31	46.36	46.41	46.47	46.52
2.70	46.57	46.62	46.67	46.72	46.77	46.82	46.87	46.91	46.96	47.00
2.80	47.05	47.09	47.14	47.18	47.23	47.27	47.31	47.35	47.40	47.44
2.90	47.48	47.52	47.56	47.59	47.63	47.67	47.71	47.74	47.78	47.81
3.00	47.85	47.88	47.92	47.95	47.99	48.02	48.05	48.08	48.11	48.14
3.10	48.17	48.20	48.23	48.25	48.28	48.31	48.34	48.37	48.39	48.42
3.20	48.45	48.48	48.50	48.53	48.55	48.58	48.60	48.63	48.65	48.68
3.30	48.70	48.72	48.74	48.77	48.79	48.81	48.83	48.85	48.87	48.89
3.40	48.91	48.93	48.95	48.96	48.98	49.00	49.02	49.04	49.05	49.07
3.50	49.09	49.11	49.12	49.14	49.15	49.17	49.18	49.20	49.21	49.23
3.60	49.24	49.25	49.27	49.28	49.30	49.31	49.32	49.33	49.35	49.36
3.70	49.37	49.38	49.39	49.41	49.42	49.43	49.44	49.45	49.46	49.47
3.80	49.48	49.49	49.50	49.51	49.52	49.53	49.54	49.55	49.55	49.56
3.90	49.57	49.58	49.59	49.59	49.60	49.61	49.62	49.63	49.63	49.64
4.00	49.65	49.66	49.66	49.67	49.67	49.68	49.69	49.69	49.70	49.70
4.10	49.71	49.72	49.72	49.73	49.73	49.74	49.75	49.75	49.76	49.76
4.20	49.77	49.77	49.78	49.78	49.79	49.79	49.79	49.80	49.80	49.81
4.30	49.81	49.81	49.82	49.82	49.83	49.83	49.83	49.84	49.84	49.85
4.40	49.85	49.85	49.86	49.86	49.87	49.87	49.87	49.87	49.88	49.88

TABLE XC — *Continued*

$\frac{x}{\text{P. E.}}$	Per Cent of Cases									
	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
4.50	49.88	49.88	49.88	49.89	49.89	49.89	49.89	49.89	49.90	49.90
4.60	49.90	49.90	49.90	49.91	49.91	49.91	49.91	49.91	49.92	49.92
4.70	49.92	49.92	49.92	49.93	49.93	49.93	49.93	49.93	49.94	49.94
4.80	49.94	49.941	49.942	49.944	49.945	49.946	49.947	49.948	49.950	49.951
4.90	49.952	49.953	49.954	49.955	49.956	49.957	49.958	49.959	49.960	49.961
5.00	49.962	49.963	49.964	49.965	49.966	49.967	49.968	49.969	49.969	49.970
5.10	49.971	49.972	49.972	49.973	49.973	49.974	49.975	49.975	49.976	49.976
5.20	49.977	49.978	49.978	49.979	49.979	49.980	49.980	49.981	49.981	49.982
5.30	49.982	49.982	49.983	49.983	49.984	49.984	49.984	49.985	49.985	49.986
5.40	49.986	49.986	49.987	49.987	49.988	49.988	49.988	49.989	49.989	49.990
5.50	49.990	49.990	49.990	49.991	49.991	49.991	49.991	49.991	49.992	49.992
5.60	49.992	49.992	49.992	49.993	49.993	49.993	49.993	49.993	49,994	49.994
5.70	49.994	49.994	49.994	49.995	49.995	49.995	49.9951	49.9952	49.9953	49.9954
5.80	49.9955	49.9956	49.9957	49.9958	49.9959	49.9960	49.9961	49.9962	49.9963	49.9964
5.90	49.9965	49.9966	49.9967	49.9968	49.9969	49.9970	49.9971	49.9972	49.9973	49.9974

APPENDIX IV

Typical Block-Design Records

The following records will illustrate the nature of the record blank used in the original standardization of the tests. Twenty designs were utilized, from which seventeen have been selected in the final revision of the test. This type of record sheet lends itself to a convenient notation of errors in performance and to recording brief notes to assist in the interpretation of the results.

NO. 26 Examined-Nov. 15, 1917-2:15 P. M.
2nd Grade
Life Age - 7y. 6m.

Wilbur C--N

1 Binet Age-8y. 0m.

16″-4

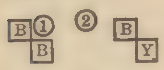
+

2 Time Limit - 11 — Seems lost- still hunting for the right block

3 Time Limit - 15 — Nothing done still trying 2 blocks - Placing flat - on side - ver. & hor.

4 Limit - 13

—



5 Limit - 12

—

Power of analysis seems lacking

6 Limit - 9

—

BLOCK-DESIGN SCORE
=3
B.D. MENTAL AGE = 6 Y. 3 M.

NO. 263 Examined-Ap. 24, 1918-10 A. M.
3-4 Grade

Albert R—S Life Age - 10y. 4m.

1 21″-7

+

2 1′19″-12

+

3 44″-9

+

BLOCK-DESIGN
SCORE = 46
B.D.M.A. =
11 Y. 10 M.

4 1′5″-11

+

5 49″-11

+

6 33″-10

+

7 2′17″-30

+

8 Limit-21

– 

9 2′8″-20

+

10 1′58″-18

+

11 Limit - 29

– 

Binet Age-
11 y. 4 m.

12 Limit - 34
(5′0″-36; not yet O.K.)

– Had O. K. once, then destroyed

13 Limit - 31

– 

14 3′54″-41

+

15 Limit - 29

– 

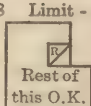
16 Limit - 49
(4′13″-52; O. K.)

–
(+)

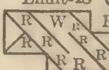
17 2′48″-30

+

18 Limit - 47

– 

19 Limit-48 (S: "The Ficture is so small & there are so many blocks, - It gits you stuck")

– 

20 Limit - 42

– 

NO. 176 Examined-Ap. 2, 1918-9 A.M.

4B Grade

Life Age - 9y. 2m.

James B—H

1 22″-8 Binet Age-8Y. 10M.

\+

NOTE:
Has set something like these at home. Larger size, but not same colors.

2 32″-4

\+

3 46″-10

\+

4 54″-11

\+

BLOCK-DESIGN SCORE = 16
B.D. MENTAL AGE = 8 Y. 9 M.

5 1′34″-15

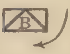
\+

6 Limit - 16

−

7 Limit - 30

−



S: "I don't see how you get these yellow ones on the side here."

8 Limit - 20

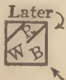
−

Later

9 Limit - 16

−

Later



10

−

SUBJECT-INDEX

AUTHOR-INDEX

U.S Department of Health, Education, and Welfare Public Health Service